TMM 3102: Protein Structure, Function and Disease

- Structural Biology Methods: Cryo-Electron Microscopy (Cryo-EM) (October 5th, 2021)

Jyh-Yeuan (Eric) Lee, Assistant Professor, BMI

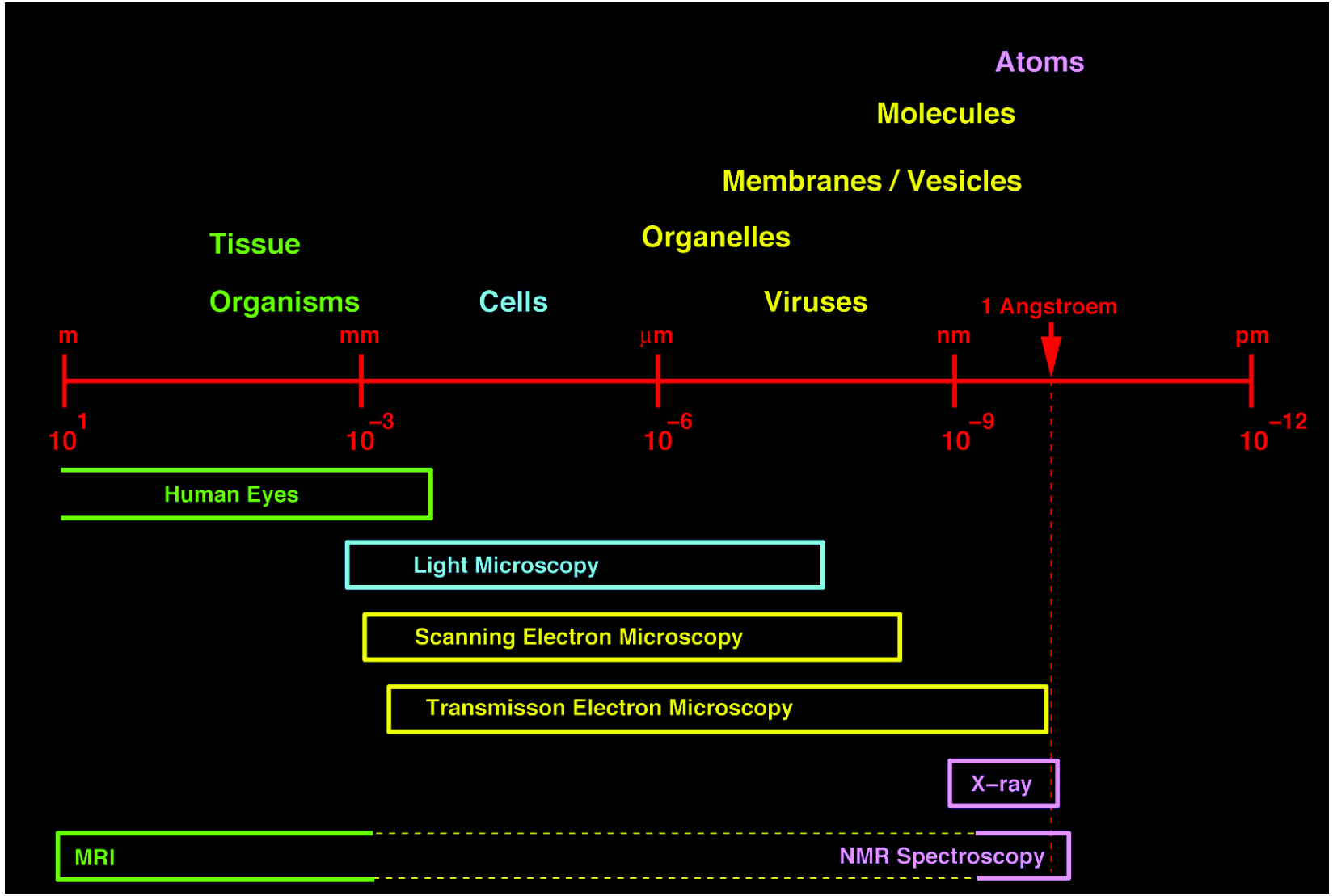
How ”Tiny” Can We See?

From humans eyes to analytical instruments, we are all limited to how small objects we can see.

For cells, we can easily observed under a light microscope and with more detailed information using electron microscopes.

To see objects at atomic resolution, so far, we know X-ray crystallography, transmission electron microscopy, and NMR spectroscopy can enable such high resolution imaging.

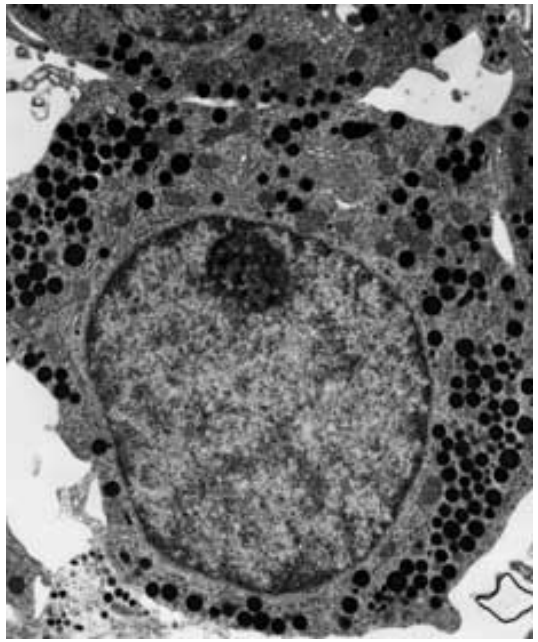
This course will selectively focuses on these three methodologies that enable vast protein structure-function studies so far.

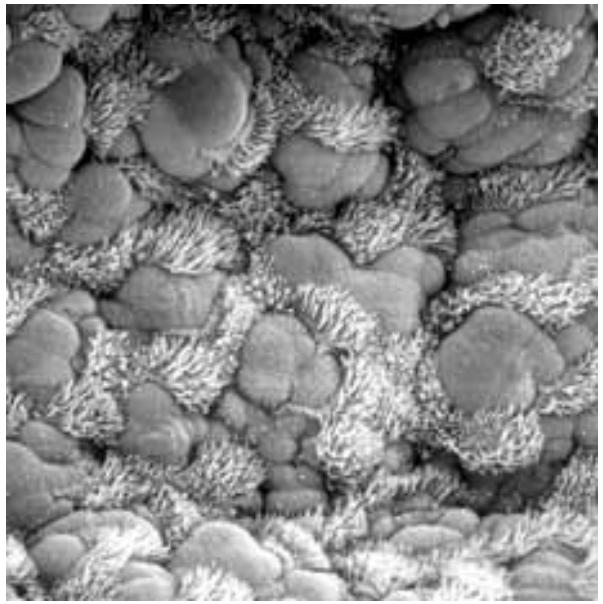


Electron Microscopy in Biology

Transmission electron microscopy (TEM)
1) section of specimen
2) 2-D projection images

Scanning electron microscopy (TEM)
1) usually entire 3-D objects
2) surface tomographic pictures





(UMass Med, EM Facility)

Travel passage of electron beams

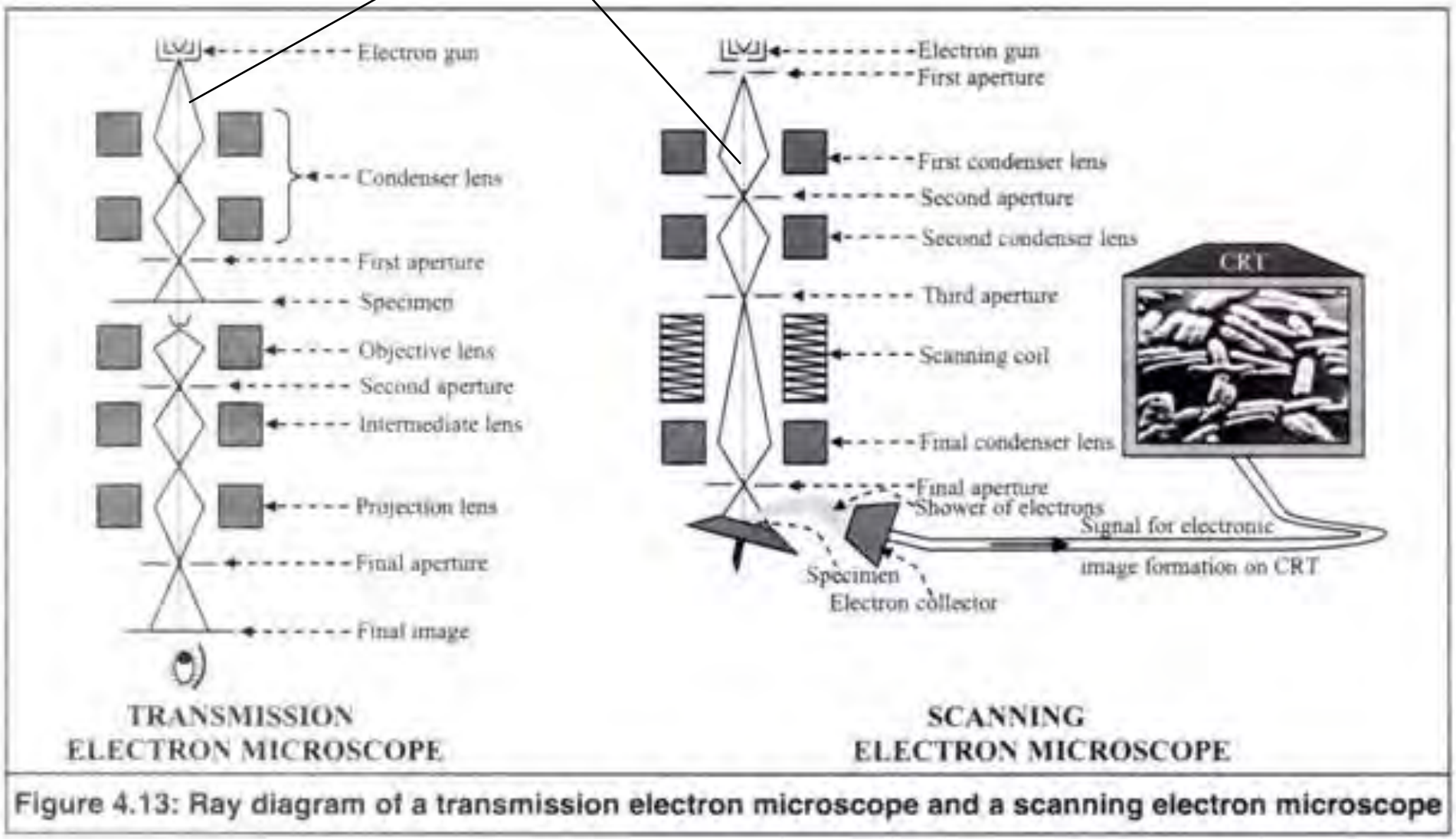
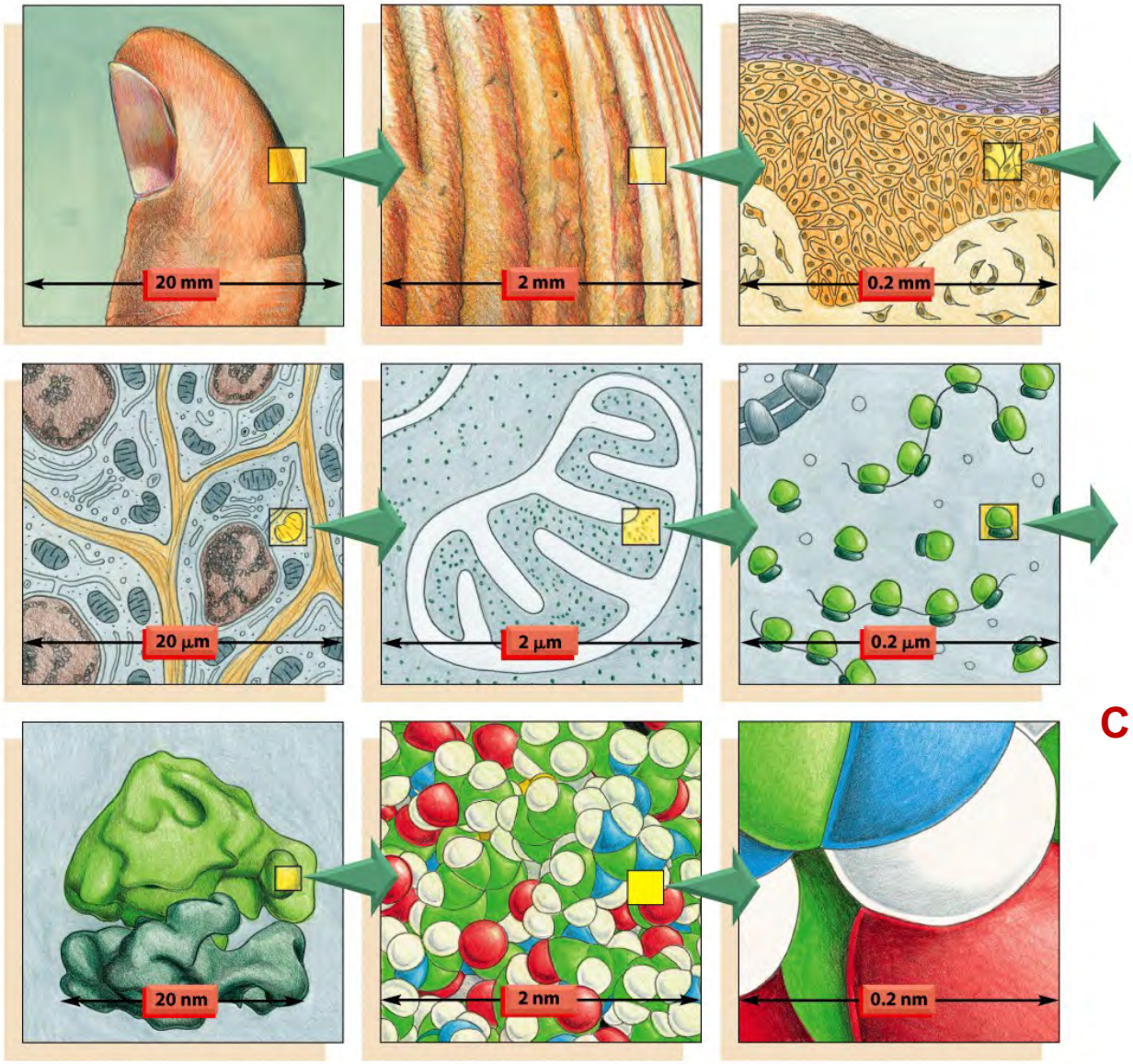
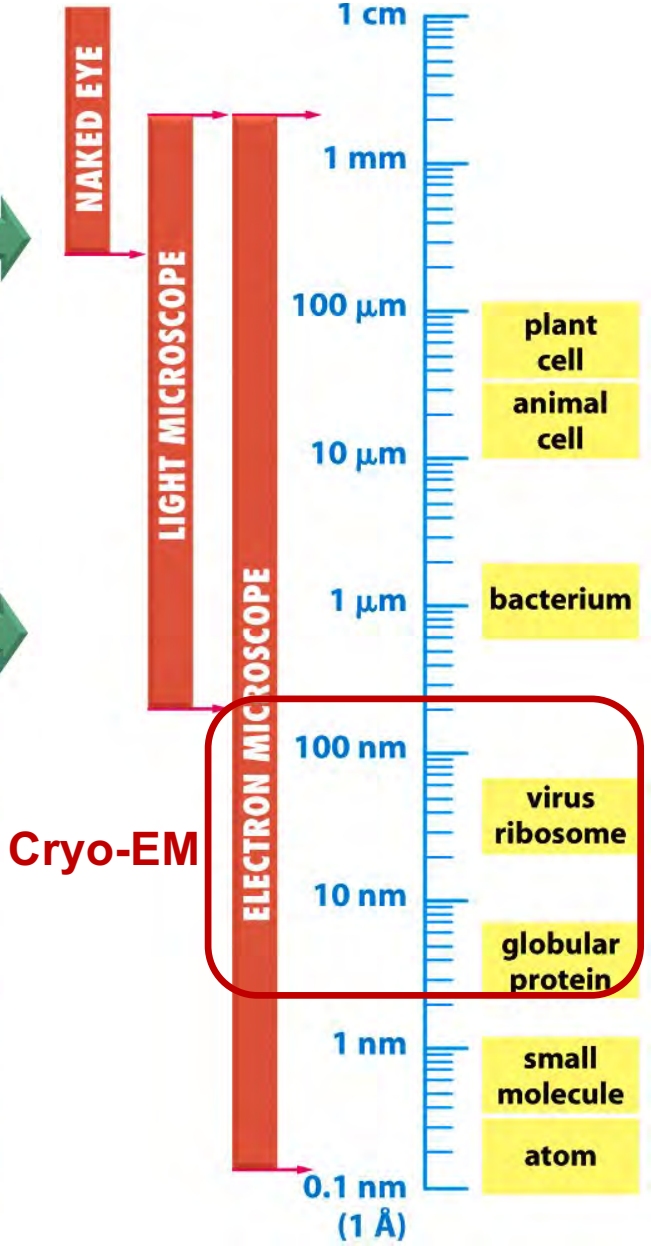


Figure 4.13: Ray diagram of a transmission electron microscope and a scanning electron microscope

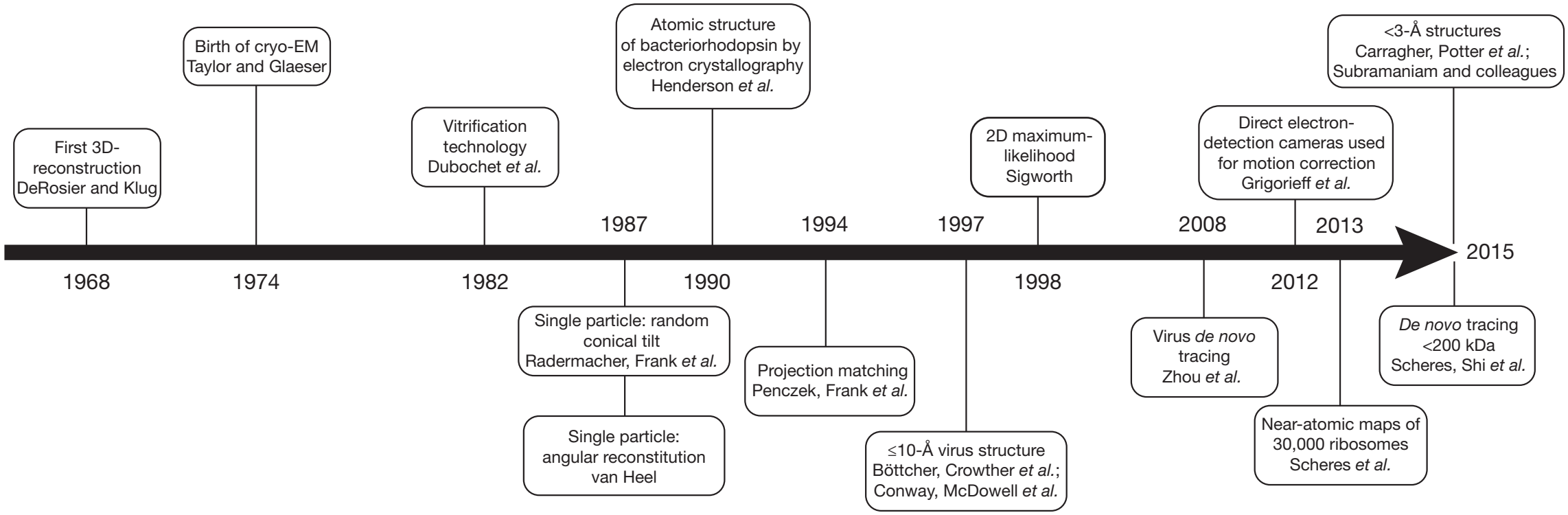
Dimensions of biological objects. Using the thumb for example, when we keep zooming-in, the dimension to see things becomes smaller and smaller:

20 mm: thumb
2 mm: fingerprint
0.2 mm: skin tissues
20 μm: skin cells
2 μm: organelles
0.2 μm: protein networks
20 nm: protein tertiary/quaternary structure
2 nm: single protein molecule
0.2 nm (or 2 Å): covalent bonds





Milestones: Timeline



(Eva Nogales, Nat Methods, 2016)

Making Structural Biology Possible

2017: Cryo-Electron Microscopy (cryo-EM)

From left: Jacques Dubochet, Joachim Frank and Richard Henderson developed cryo-electron microscopy.

AWARDS

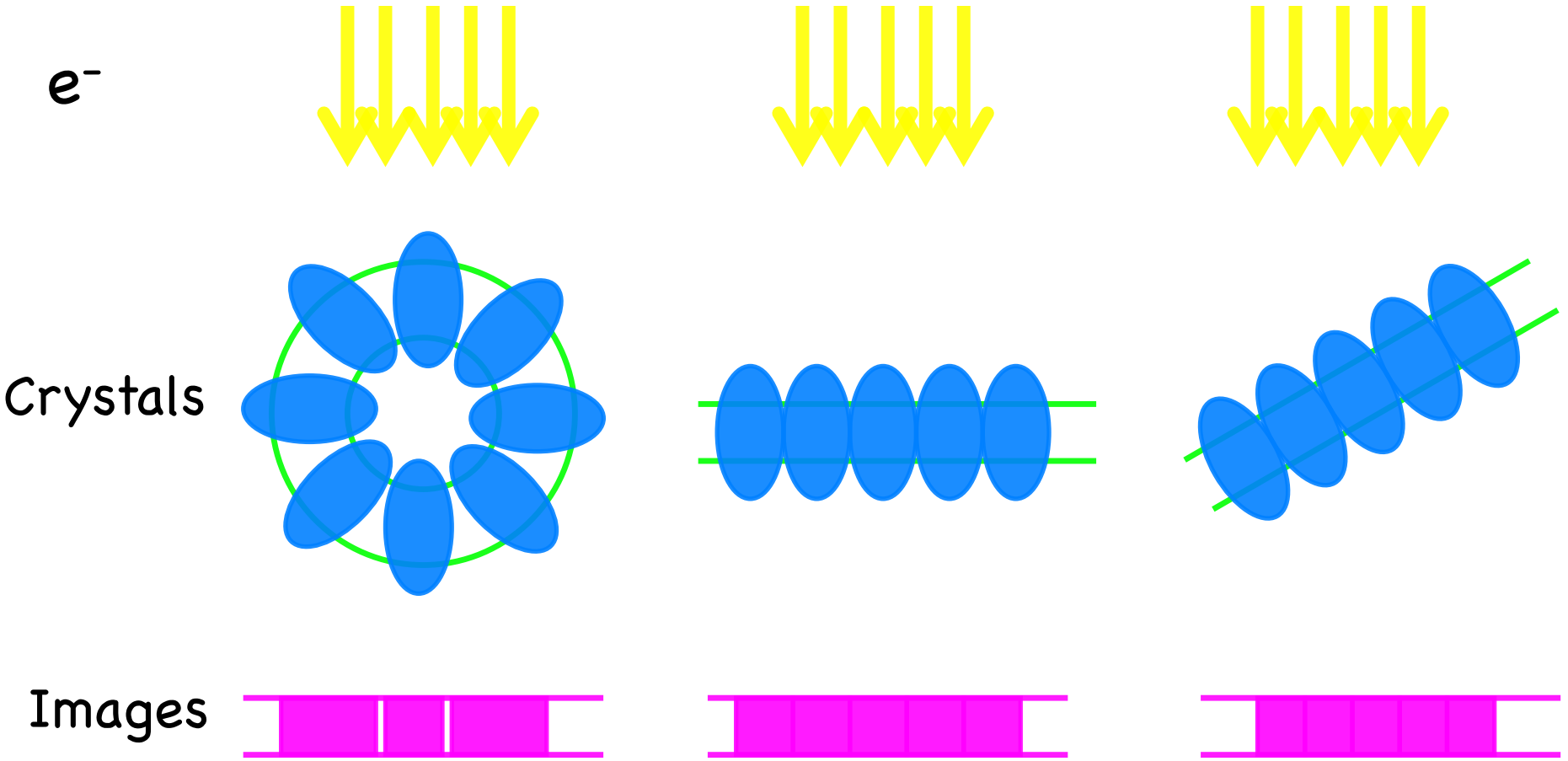
Molecular-imaging pioneers scoop Nobel

Chemistry prize hails work on cryo-electron microscopy.

(Cressey & Callaway, Nature, 2017)

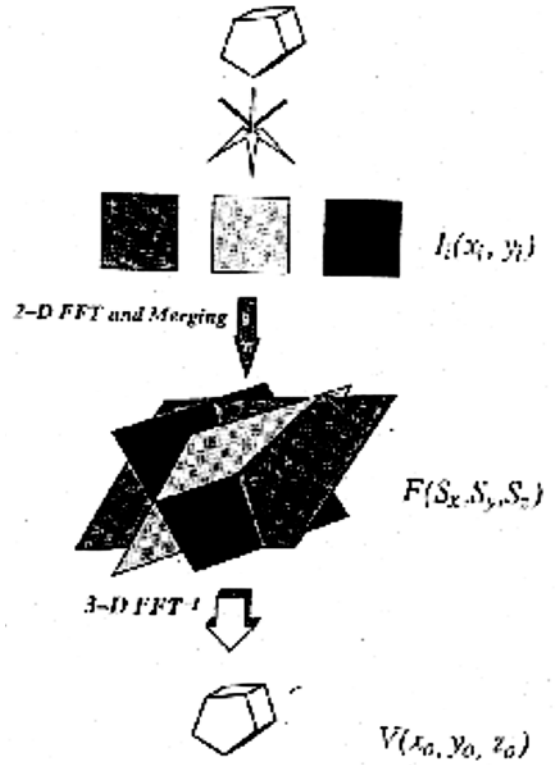
Cryo-EM: Method of the Year 2015

Only 2-D projections are recorded (x,y)!!



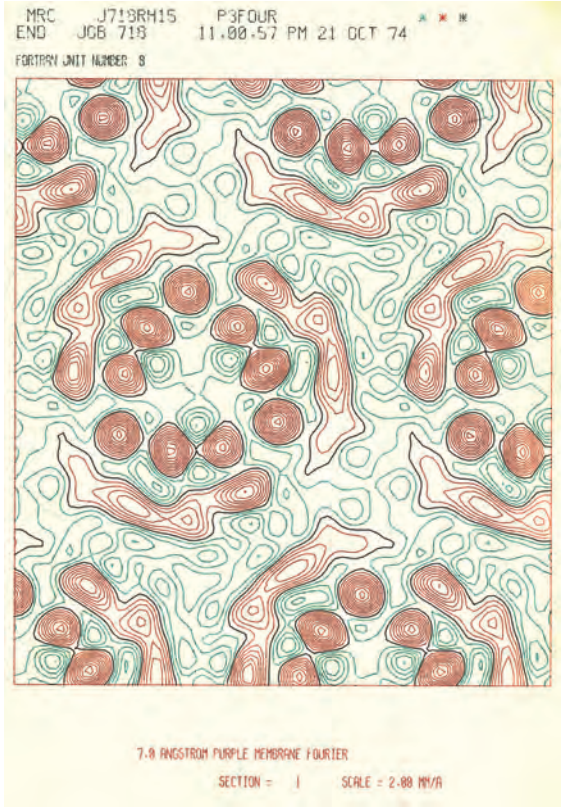
Milestones: 3-D Reconstruction

Principle



(Amos et al, Prog Biophys Mol Biol, 1983)

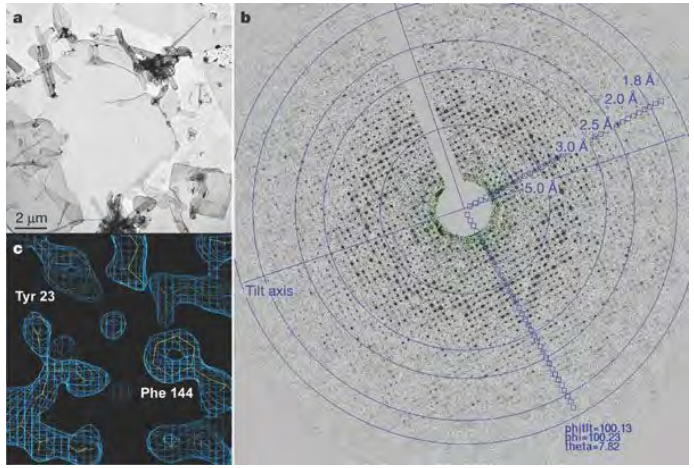
First Report (7 Å, 1974)



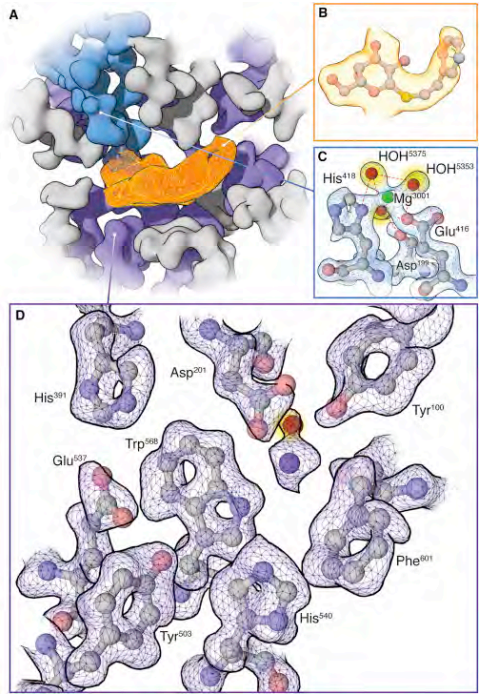
(Unwin & Henderson, J Mol Biol, 1975)

Milestones: 3-D Reconstruction

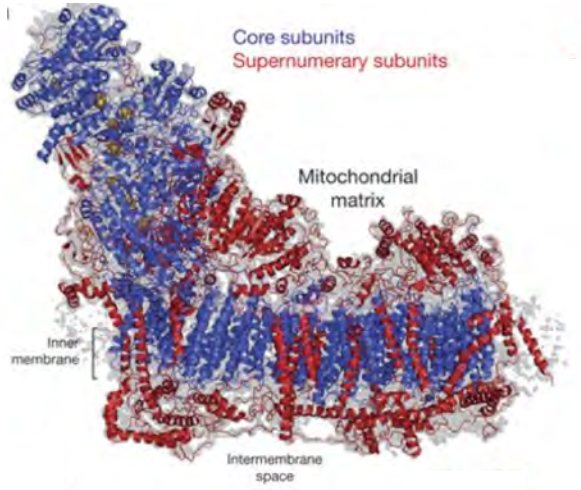
Higher Resolution (~2Å) and Bigger Molecules/Complexes



Aquaporin (1.9 Å)
(Gonen et al, Nature, 2005)



β-Galactosidase (1.5 Å)
(Bartesaghi et al, Structure, 2018)



Mitochondrial Complex I (4.2 Å)
(Zhu et al, Nature, 2016)

October surprise: 1.22/1.25 Å (apoferritin)

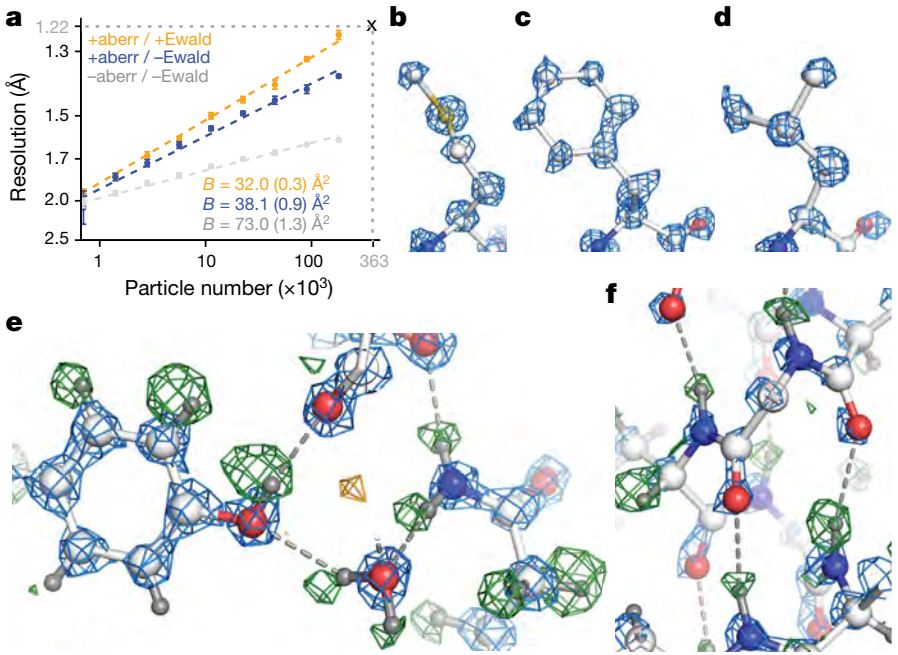


Fig. 3 | Apoferritin reconstruction. a, *B*-factor plots for reconstructions using: high-order aberration (+aberr) and Ewald sphere correction (+Ewald; orange); high-order aberration correction only (blue); and no correction (grey). *B*-factors estimated from the slopes of fitted straight lines are shown in the same colours. Numbers in parentheses and error bars represent estimated and sample s.d.s from sevenfold random resampling, respectively. **b**–**d**, Densities from the 1.22 Å map (blue) for M100 (**b**), F51 (**c**) and L175 (**d**). **e**, Hydrogen-bonding network around Y32 and water-302 is visible in the difference map (green, positive; orange, negative). **f**, The α-helix hydrogen-bonding network involving residues ^{21}NRQIN^{25}, shown as in **e**.

(Nakane et al, 2020, Nature)

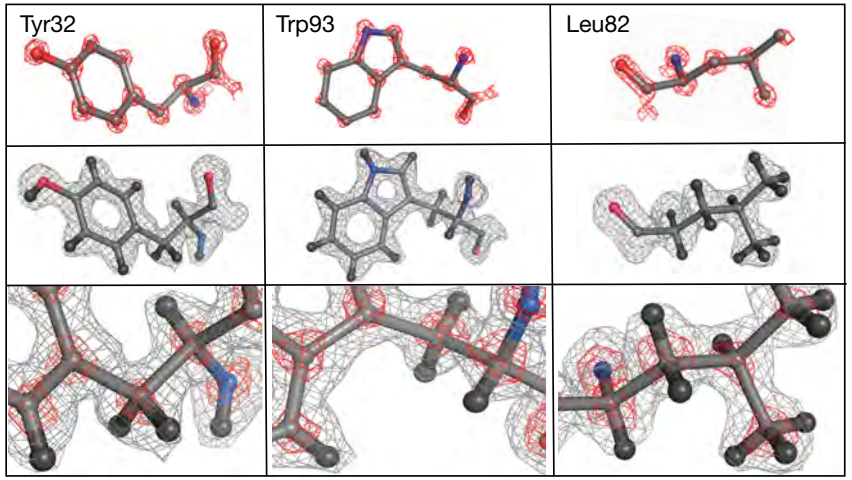


Fig. 3 | True atomic resolution. Visualization of individual atoms and hydrogens at 1.25 Å resolution. Three apoferritin residues are shown at high (red mesh) and low (grey mesh) density thresholds. The true atomic resolution of our map is shown in the first row by the clear separation of individual C, N and O atoms at high thresholds. The second row shows density that agrees with hydrogen atom positions in all parts of the individual amino acid side chains. A ball and stick representation for the hydrogen atoms (black) is included in the atomic model. Similar visibility of density for hydrogen atoms requires about 1 Å or better resolution in X-ray crystallographic structures (Extended Data Fig. 3). The third row shows close-up views of the three amino acids using both density thresholds simultaneously.

(Yip et al, 2020, Nature)

Single-Particle Analysis (SPA)

No Need of Protein Crystals!

Cryo-EM applications in protein structural biology:

1) Single particle analysis:
(no need to generate protein crystals)

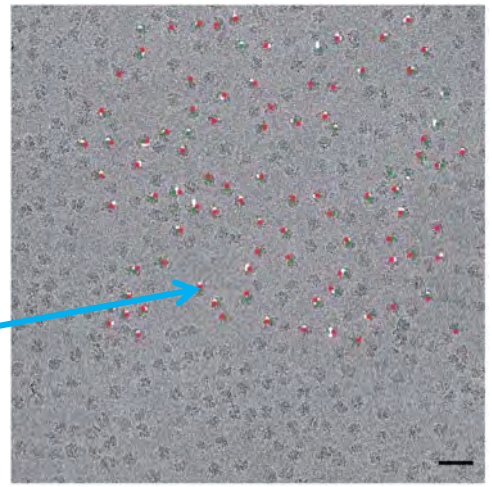
a. Prepare purified protein samples and freeze them.

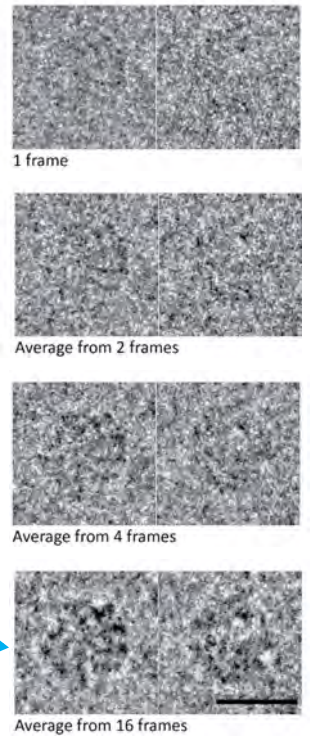
b. Take electron micrographs of isolated protein particles.

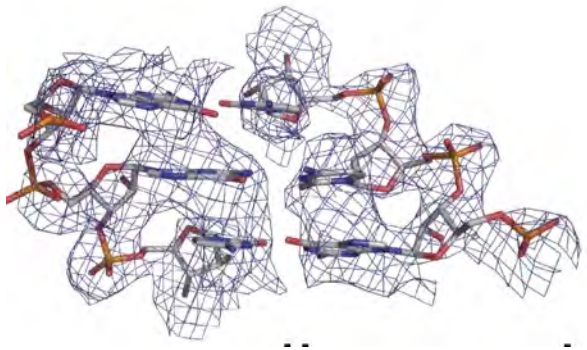
c. Collect several of “identical” images and add altogether to enhance the signal of the objects (darker area)

d. Then place amino acid models in the electron density.

Samples







(Bai et al, eLife, 2013)

Microelectron Diffraction (MicroED) & Electron Crystallography

Cryo-EM applications in protein structural biology:

2) Microelectron diffraction:
(need to generate protein crystals, but much smaller than that for X-ray crystallography)

a. Prepare purified protein samples, crystallize and and freeze them.

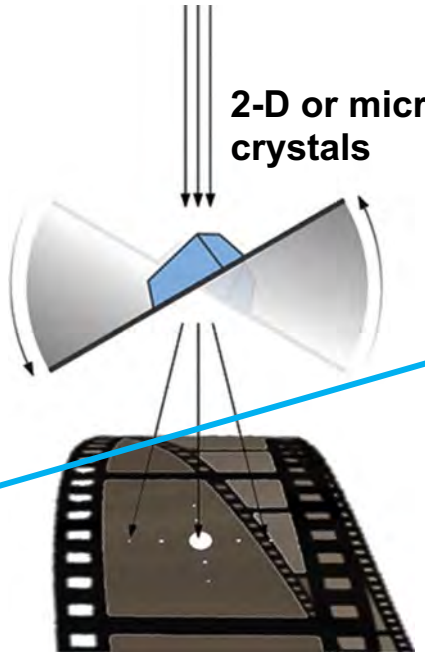
b. Take electron micrographs of protein crystals and diffraction images.

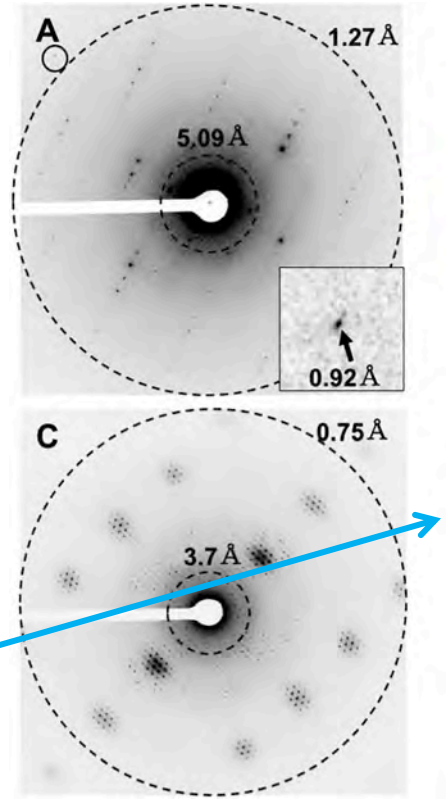
c. Use diffractions to establish the amplitude information and images to find the phase information, then calculate the structural factors to generate the electron clouds for the target proteins.

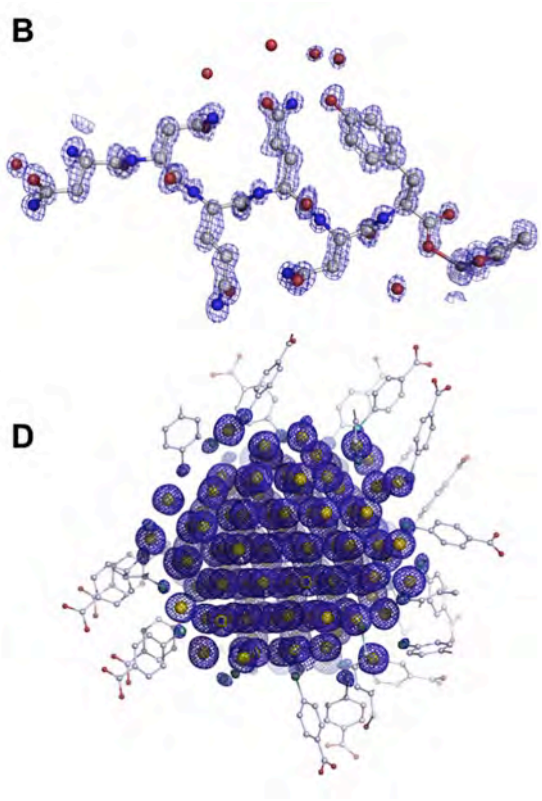
d. Model building like that in X-ray crystallography.

Samples

2-D or micro-crystals







(Martynowycz & Gonen, Curr Opin Colloid Interf Sci, 2018)

Cryo-Electron Tomography (cryo-ET)

Cryo-EM applications in protein structural biology:

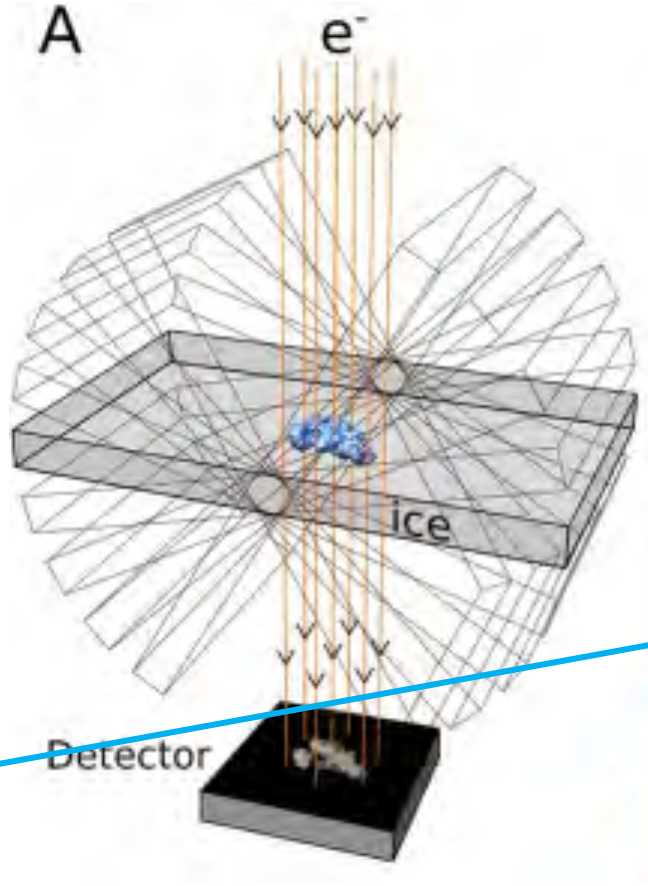
3) Electron tomography:
(suitable for large protein-protein network, organelles, subcellular structures)

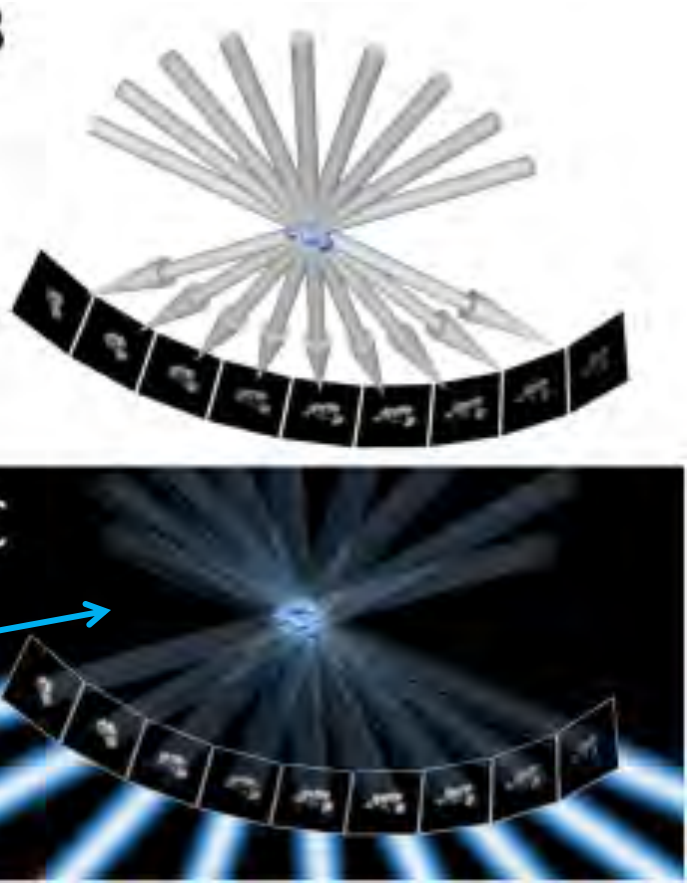
a. Prepare purified/ homogenous samples and freeze them.

b. Take electron micrographs of isolated protein particles at different angles almost at the same time.

c. Reconstruct the 3-D images instantaneously.

Samples





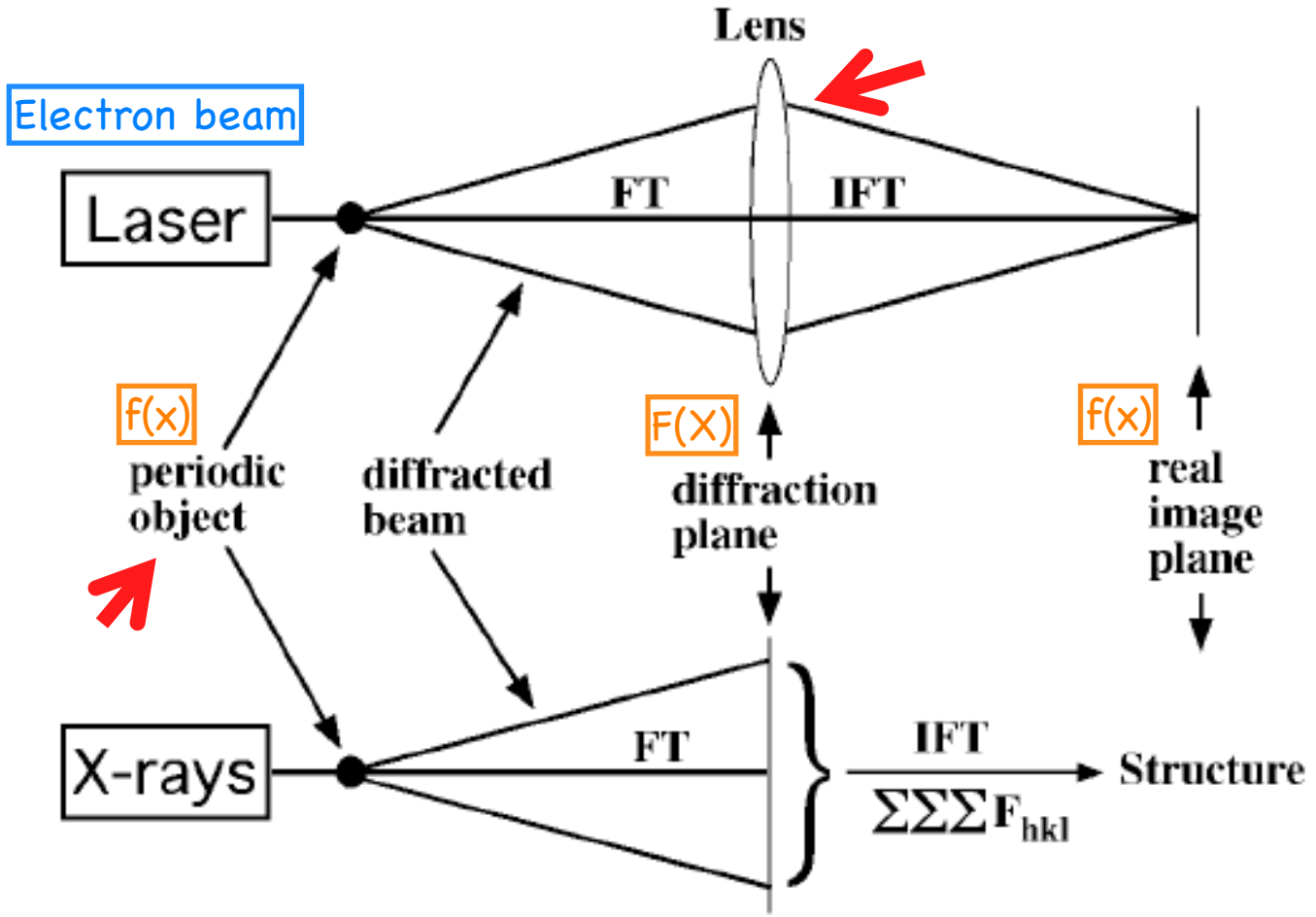
(Koning et al, Ann Anat, 2018)

Optical & X-ray Diffractions

Schematic diagrams to illustrate the optical difference between X-ray crystallography (bottom) and electron/light microscopy (top):

Microscopy (top):
When a periodic object is subjected to laser or electron beams, the diffracted beam (information in reciprocal space, i.e., FT product) is processed by a physical len which then converts the information back to image in real space, i.e., IFT.

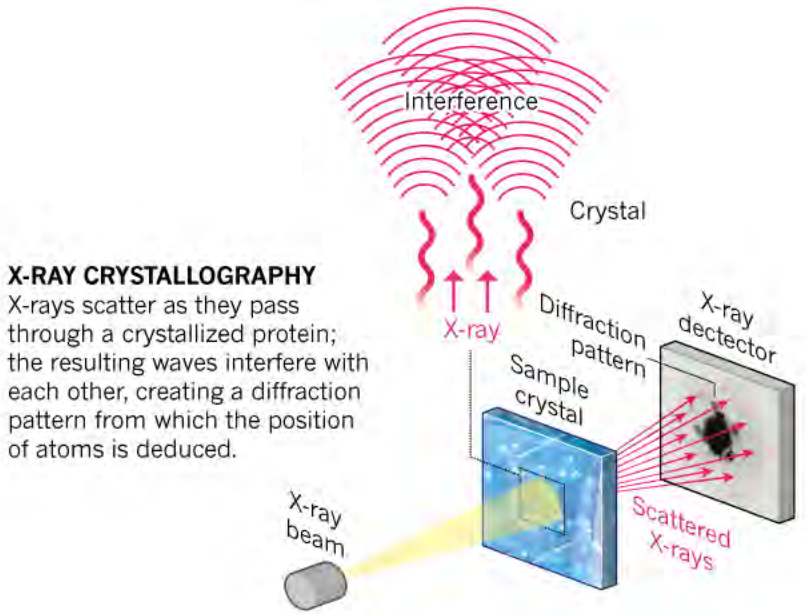
X-ray (bottom):
The information of the periodic object is only collected on the diffraction plane, because of lack of X-ray lens. To generate the real image back, it requires algorithms to process such IFT operation as would be done in microscopy.



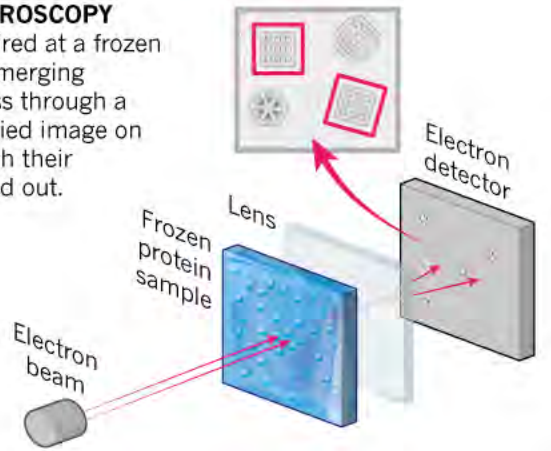
If F(X)=FT[f(x)], then f(x)=IFT[F(X)], where FT=Fourier transform & IFT=Inverse Fourier transform.
Note: important in X-ray crystallography and 3D reconstruction algorisms.

Optical & X-ray Diffractions

Schematic diagrams to illustrate the instrumental difference between X-ray crystallography (left) and cryo-electron microscopy (right)

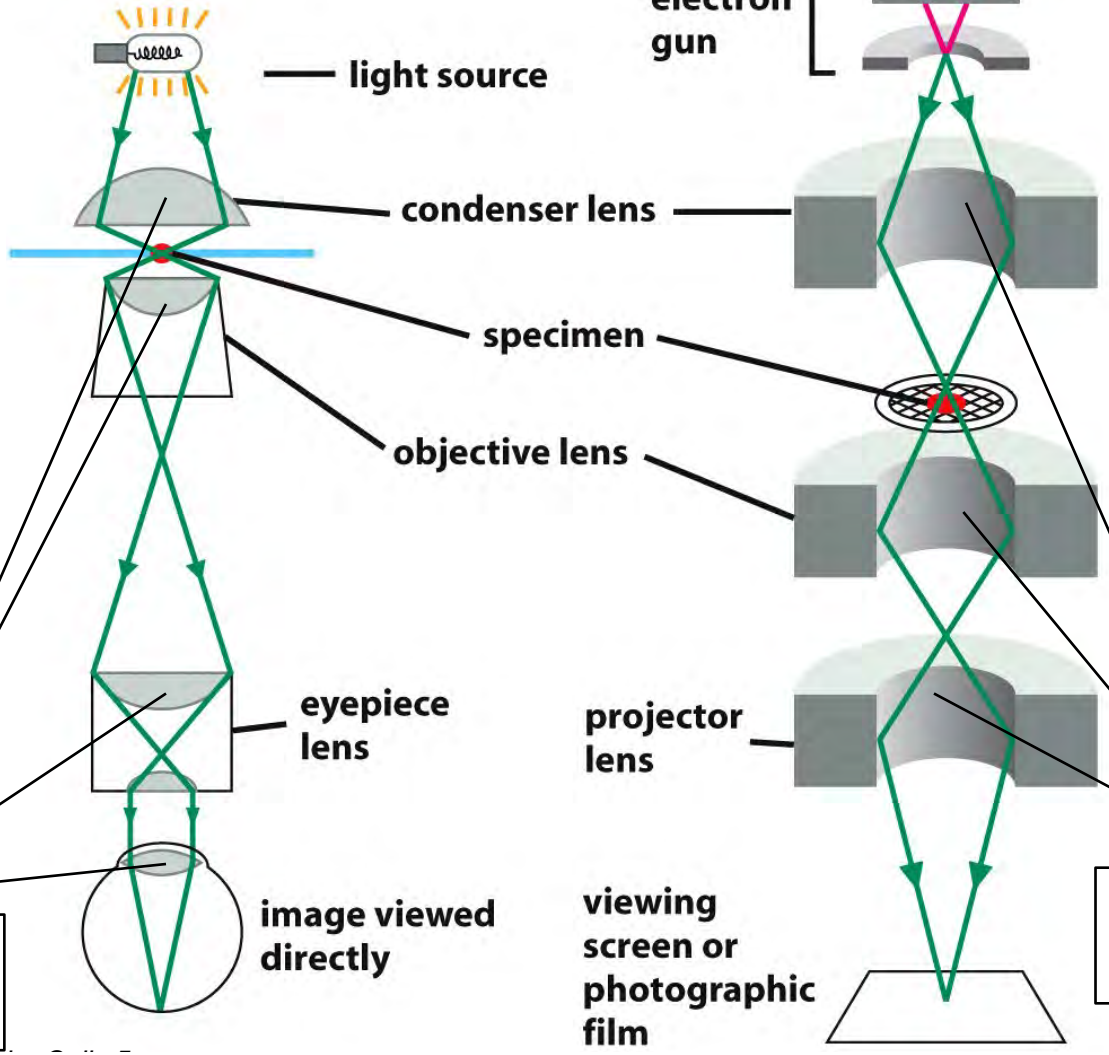


X-RAY CRYSTALLOGRAPHY
X-rays scatter as they pass through a crystallized protein; the resulting waves interfere with each other, creating a diffraction pattern from which the position of atoms is deduced.



CRYO-ELECTRON MICROSCOPY
A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.

Anatomy of light and electron microscopes



Light microscope:
1) bright-field
2) fluorescence
3) UV

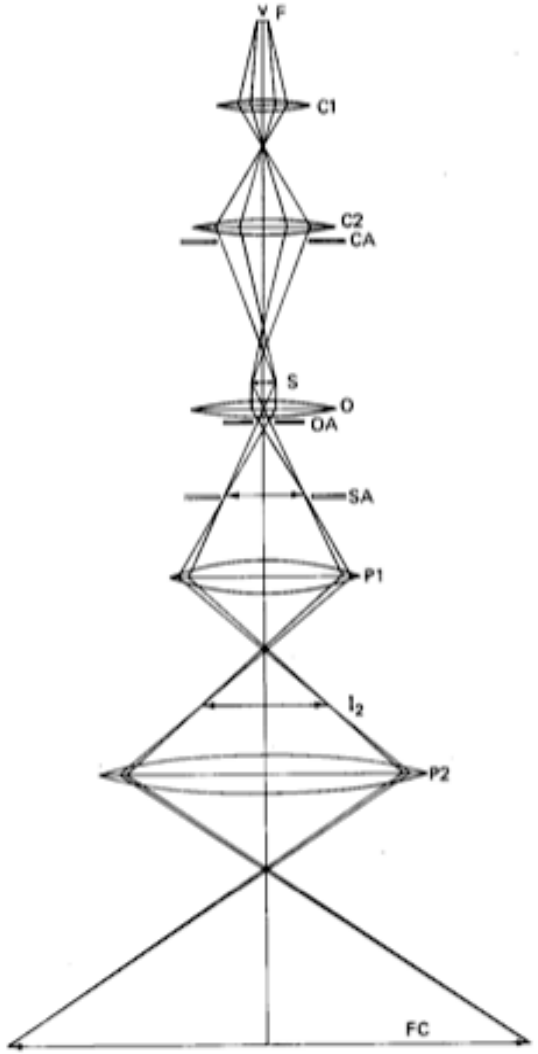
Usually made of glass/quartz

Electron microscope:
1) transmission (TEM)
2) scanning (SEM)
3) scanning/transmission (STEM)

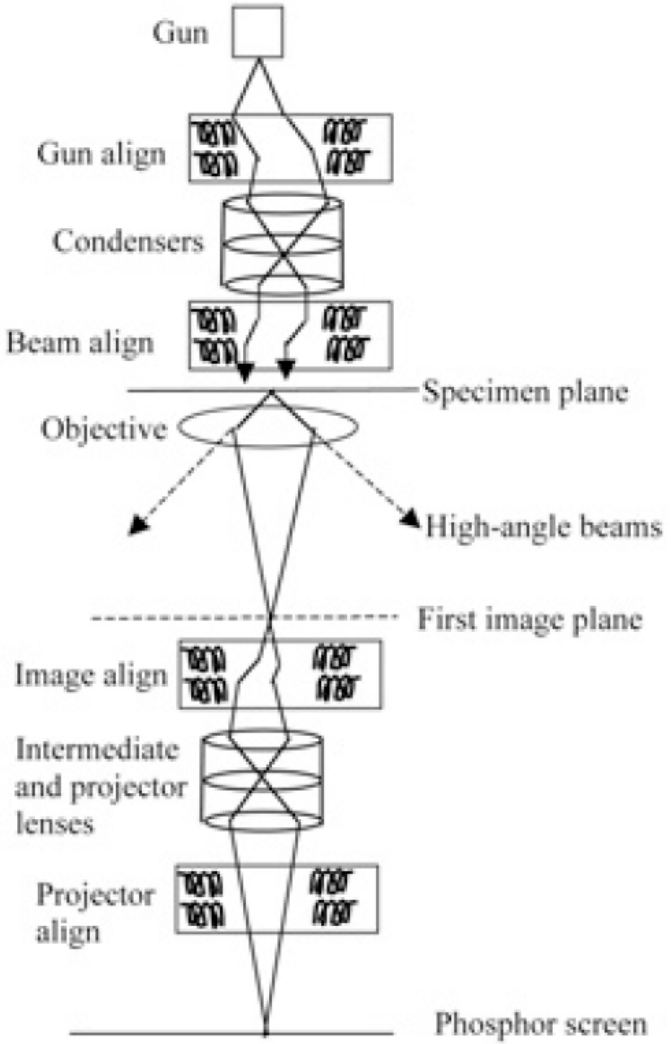
Made of electric magnets

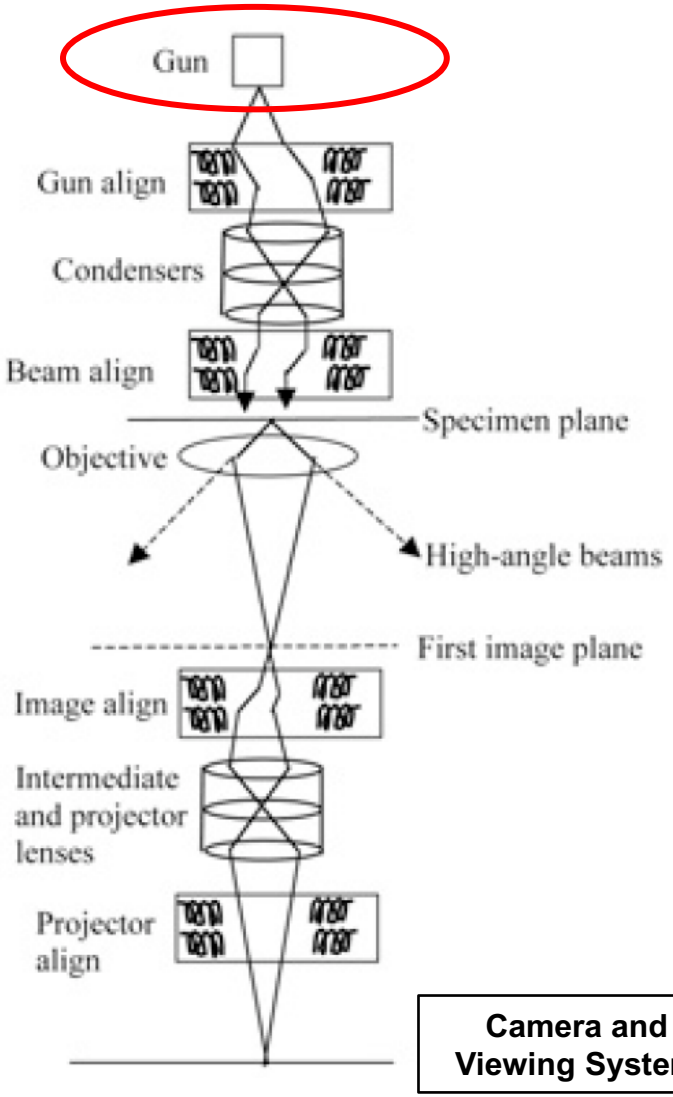
Why do we need to know optics?

Ideal instrument

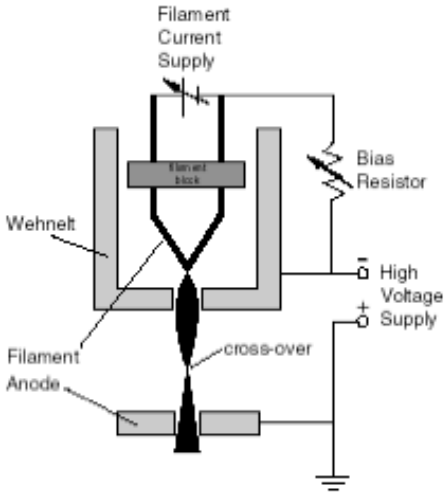


Real instrument



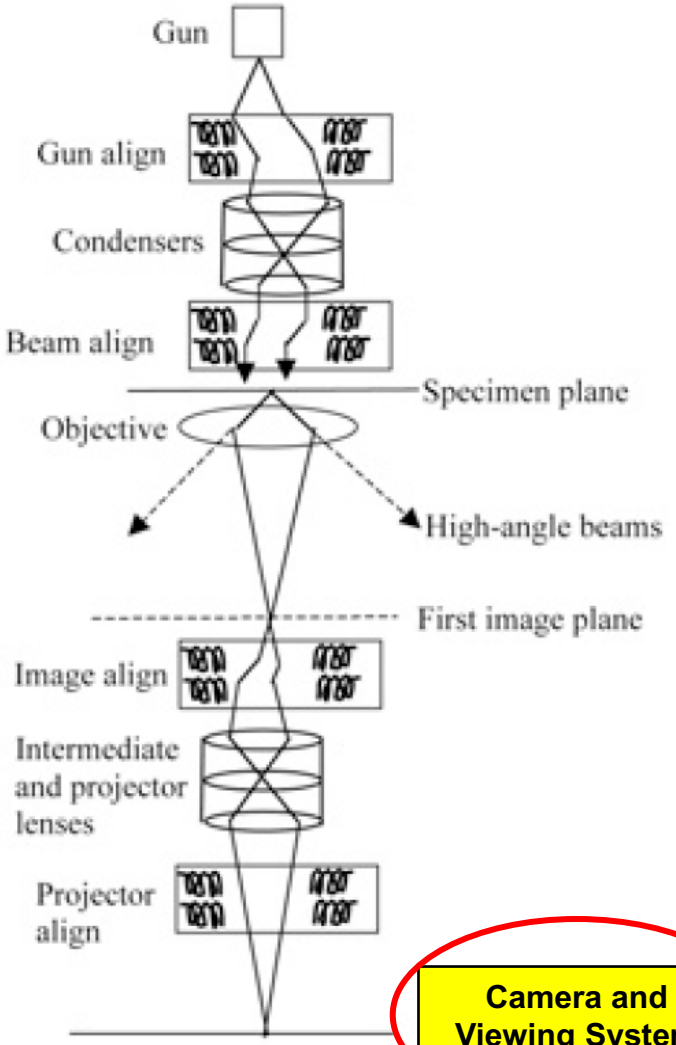


Electron Gun

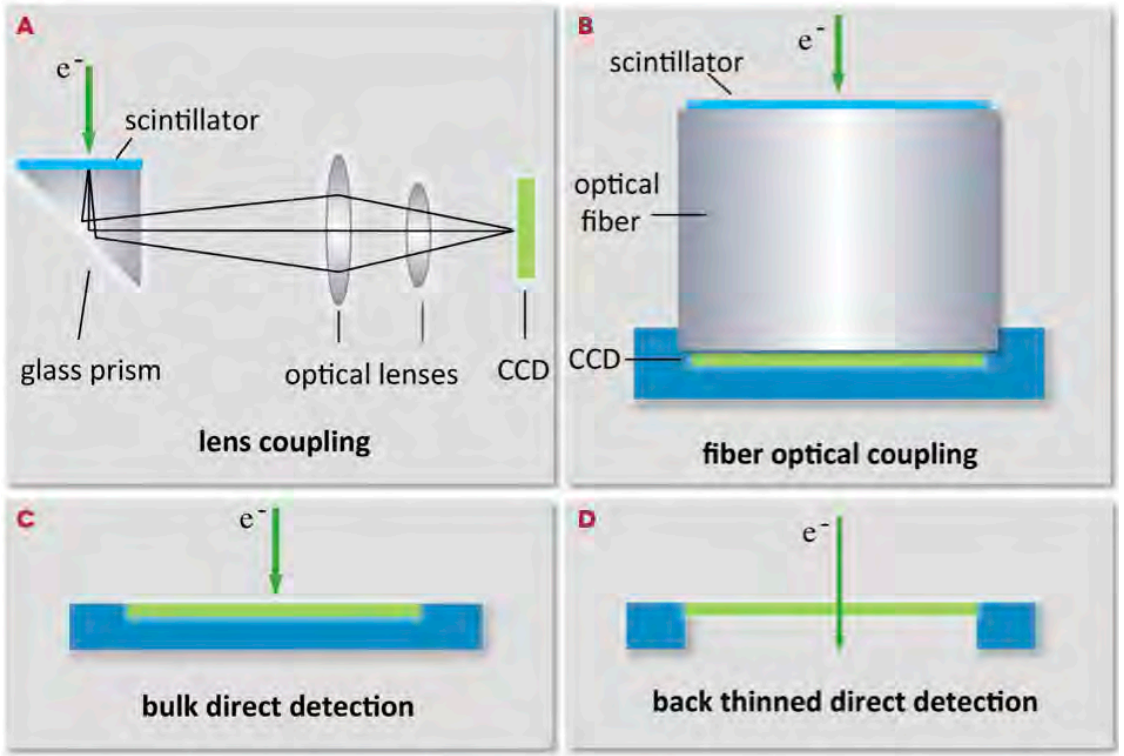


- Thermionic Emission Gun
  - Electrons are emitted from the heated filament.
  - Lanthanum Hexaboride (LaB6) Single crystal: ~1900K
  - Hairpin Tungsten (W) filament: ~2600K

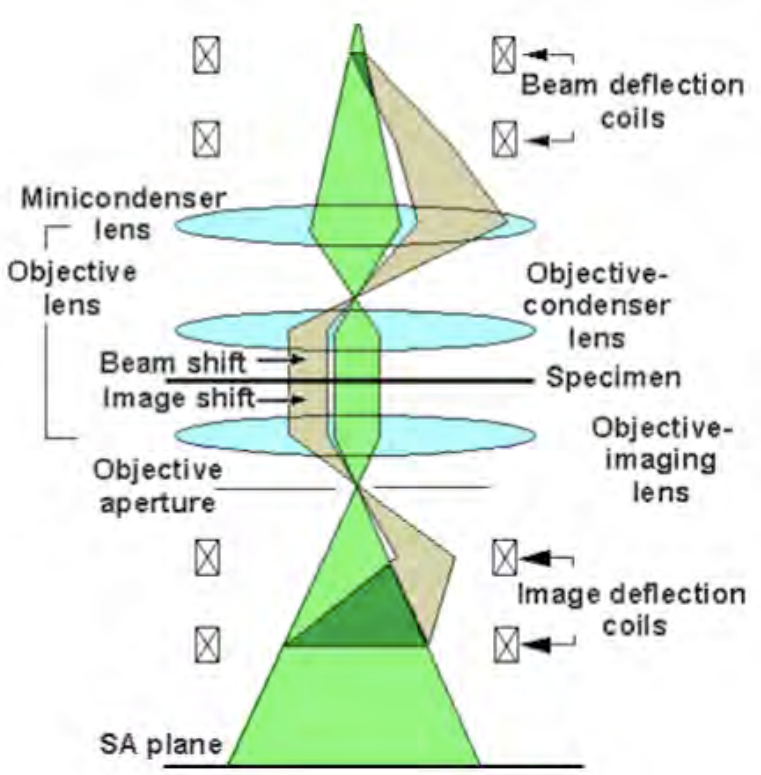
- Field Emission Gun **(high coherency)**
  - Electrons are extracted from the W emitter tip surface by the high electric field ~10V/nm.
  - Cold cathode type: room temperature
  - Schottky type: ~1800K

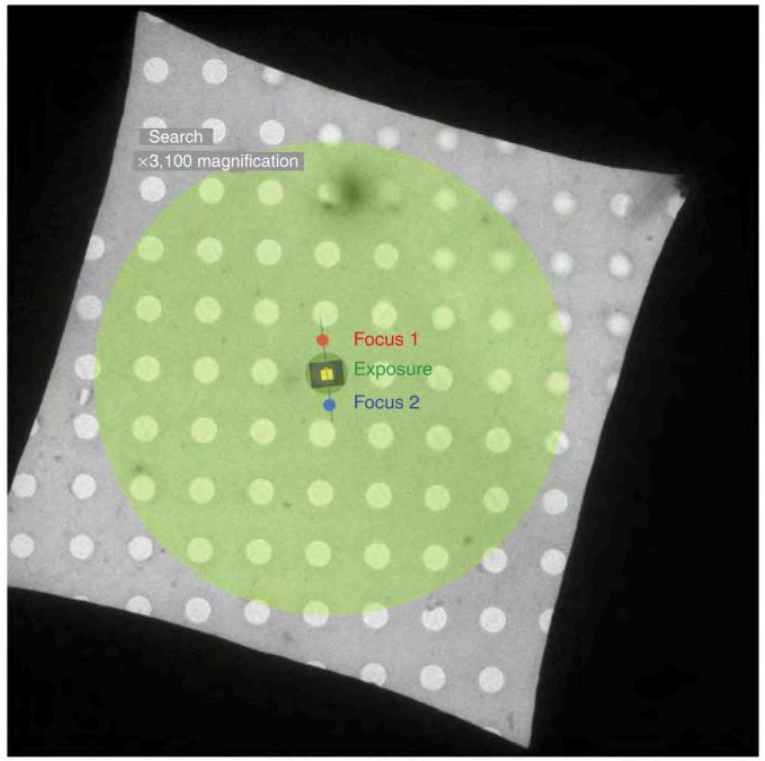


Phosphor Screen or Digital Detectors



Low-Dose Imaging





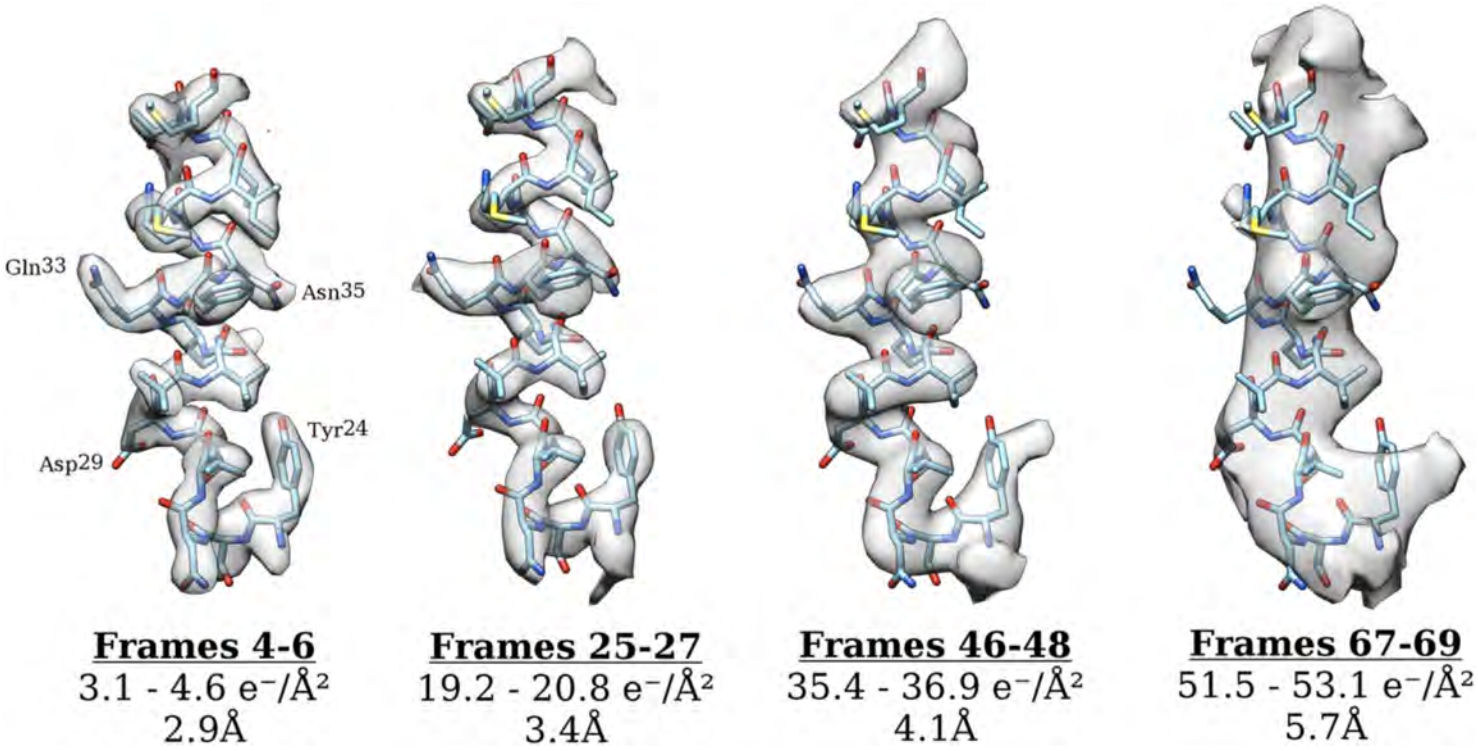
Grassucci … Frank. Nat Protoc 2008;3(2):330–339

- Search Mode: Low mag, low dose rate 10^{-3} e/Å^2/sec
- Focus Mode: High dose in an adjacent area at chosen magnification
- Photo Mode: Defocus at the desirable value for an anticipated resolution. Use a total dose of ~20-50 e/Å^2

Defocus will be an important factor to determine the image quality.

Imaging: Parameters to Consider

- Dose rate: “low-dose” mode



(Grant & Grigorieff, eLife, 2015)

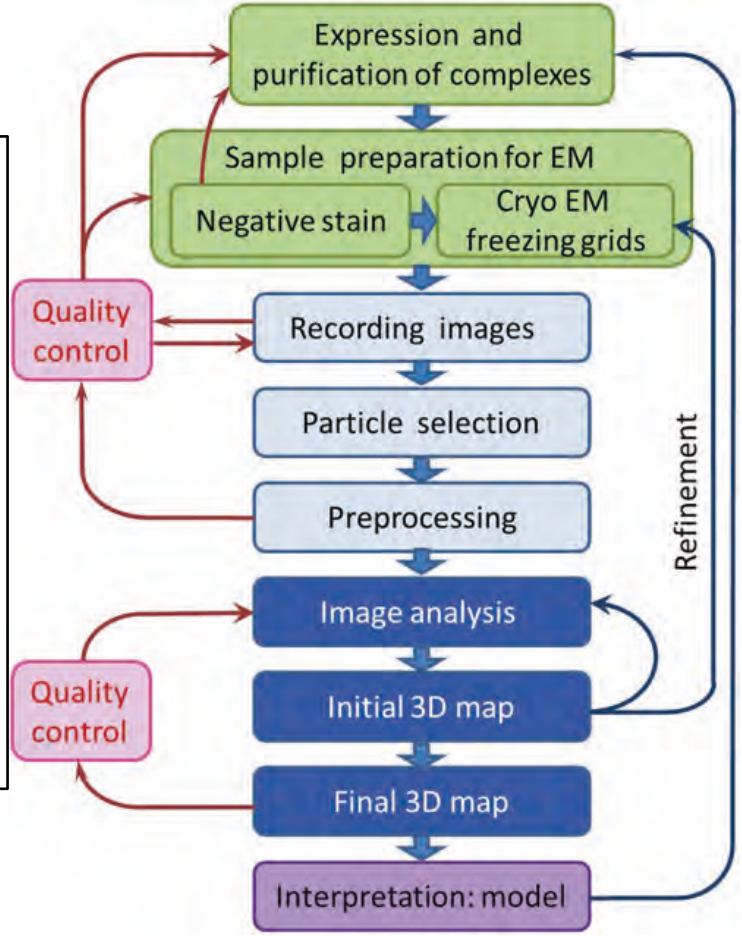
Cryo-EM Workflow

Green:
Protein biochemistry
Sample preparation

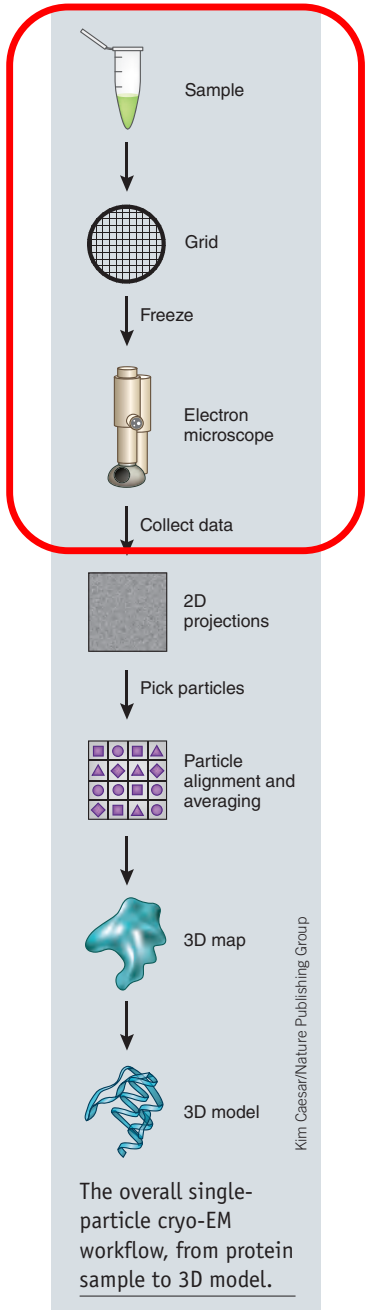
Light blue
Imaging
Data collection

Blue
Image processing
Model building

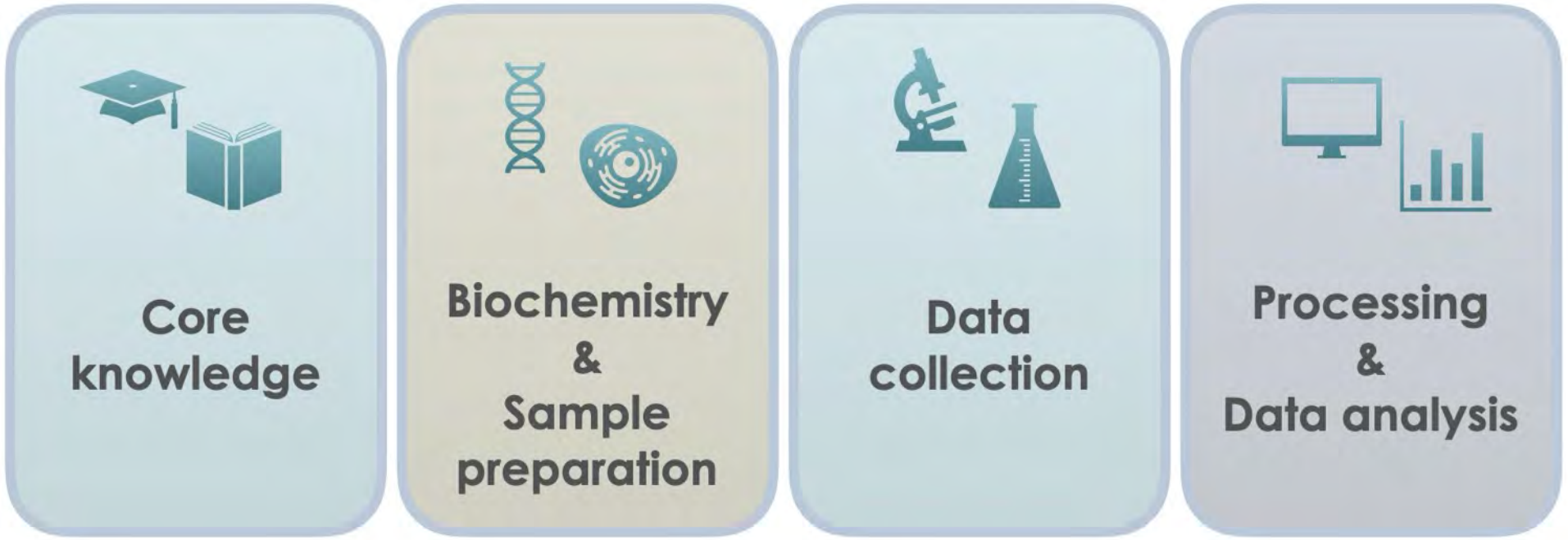
Purple
Model validation



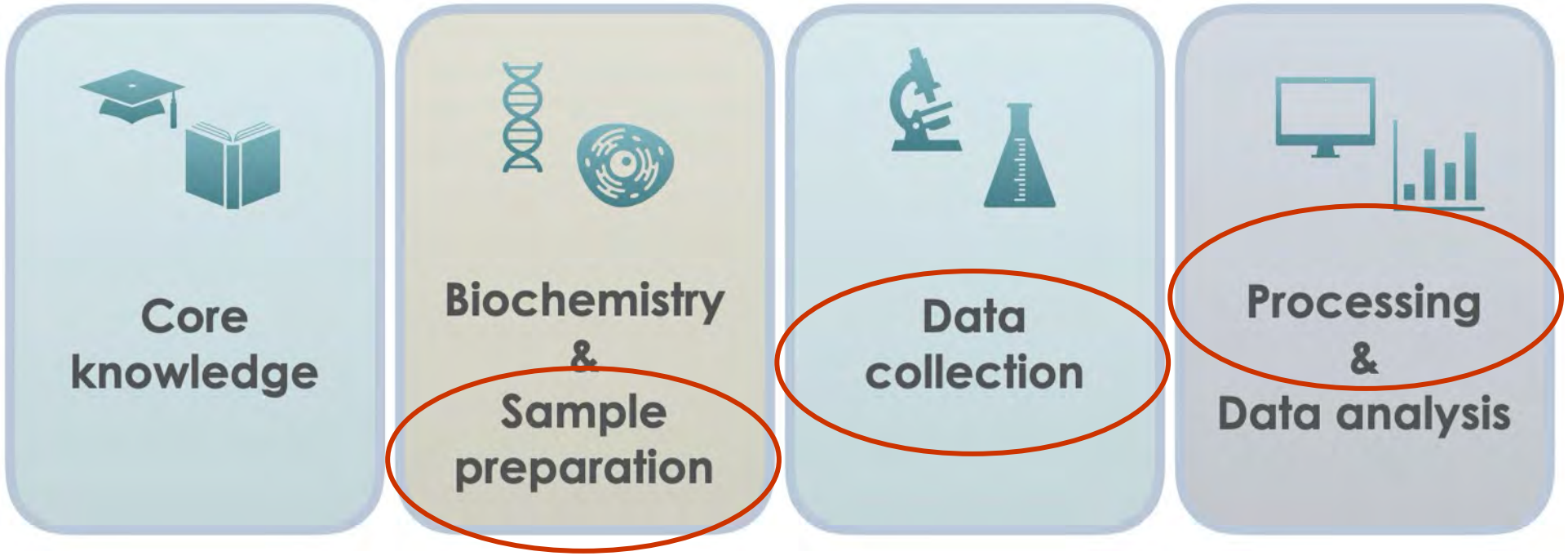
(Costa et al, Meth Mol Biol, 2017)
(Dörr, Nat Meth, 2016)



How do we start? 4 aspects.



Bottlenecks



Protein Sample Quality

So what kind of sample quality do we need to do cryo-EM?

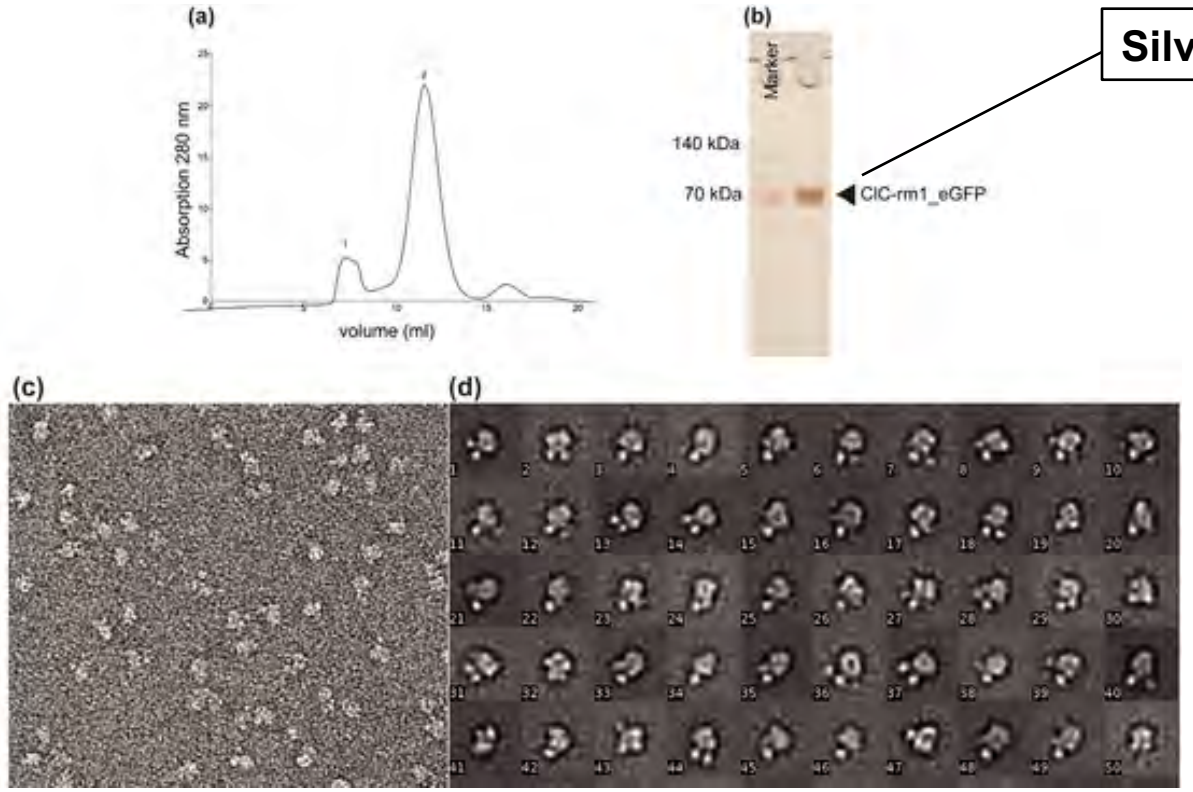
Use single-particle analysis as example:

a. It needs to be as pure as that used for crystallization.

b. No need of mg quantity of protein yields. In general, when one can see a clean and sharp protein bands by silver staining, it's sufficient enough.

Note: After running an SDS-PAGE gels, proteins can be stained in many ways. Common staining is done with Coomassie blue dye, but its sensitivity is limited to **µg-mg** proteins. Silver staining however can detect protein quantity at as low as **ng**.

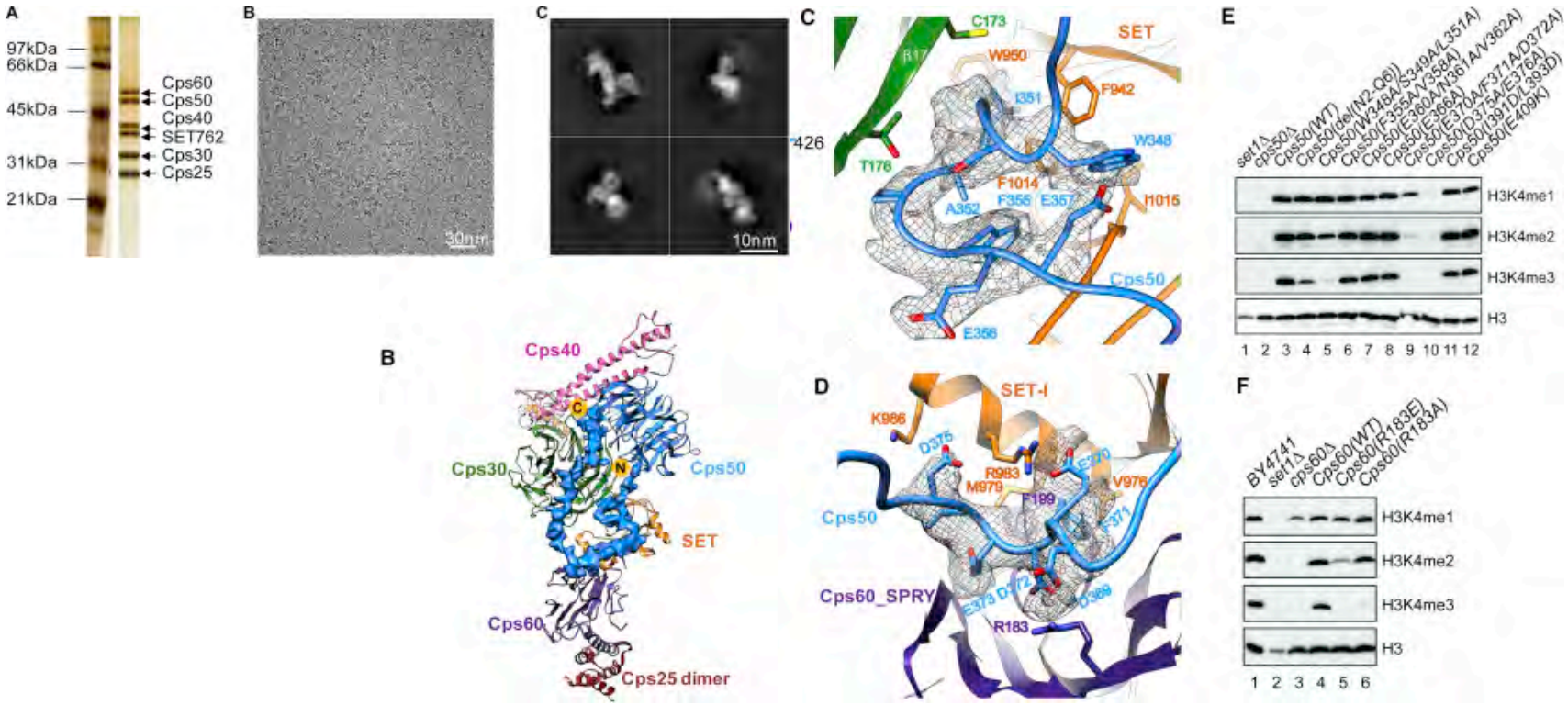
(But no need to purify several mg proteins!)



(Abeyrathne & Grigorieff, PLOS ONE, 2017)

Protein Sample Quality

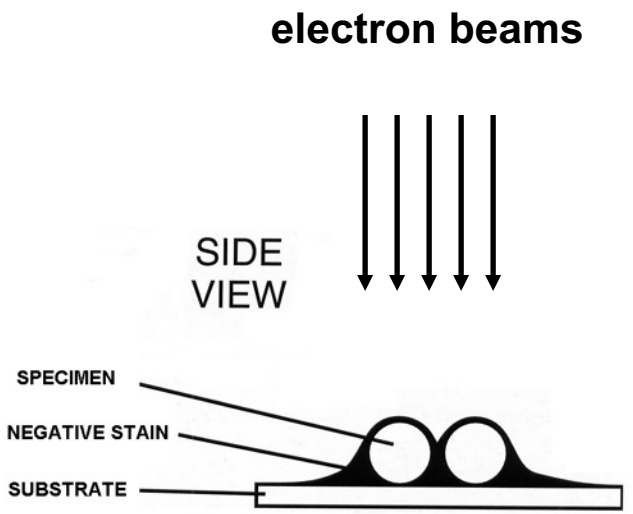
(Functionally characterized)



(Qu et al, Cell, 2018)

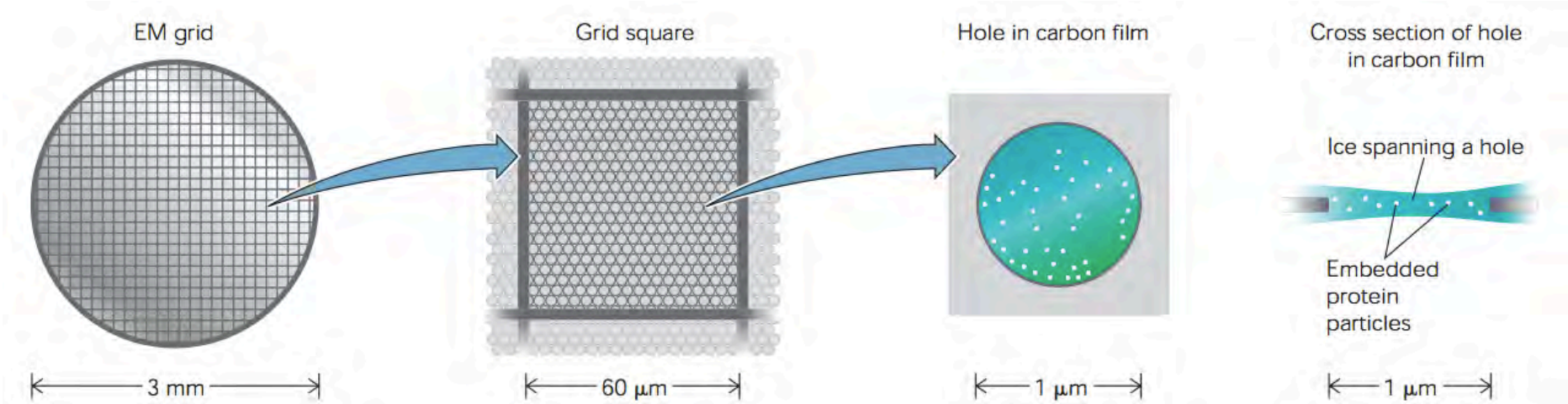
EM Samples: Negative Staining

- Principle:
  - Embedding objects in a layer of heavy-metal salts that surround the proteins like a shell.
  - Shape of objects are visible in contrast to the optically opaque stains.
- Benefits:
  - Small amount of proteins (0.01 mg/mL)
  - Easy and quick (preparation and imaging)
  - No need of high-end microscope; diagnostic
- Downsides:
  - Low resolution (*e.g.*, high noise from stains)
  - Artifacts (lack of hydration)



(Brenner & Horne, BBA, 1959)

Preparation of cryo-EM Grids

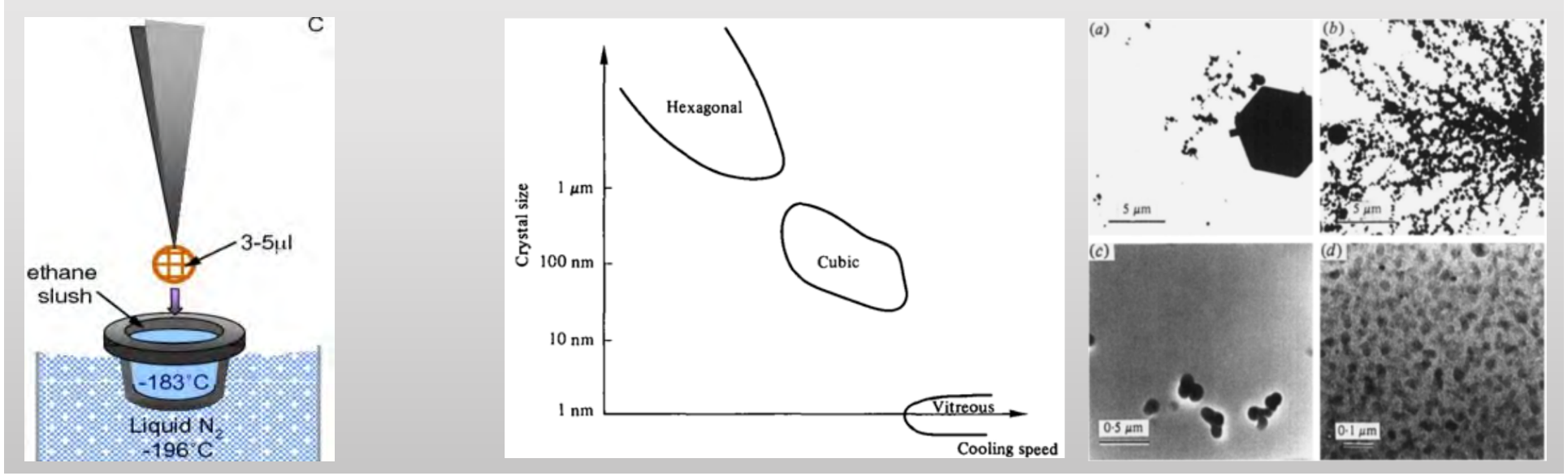


The protein solution is spread on a very thin carbon film perforated with small 1 μm diameter holes. The sample is rapidly frozen in liquid ethane to form vitreous ice (water molecules frozen randomly, not in an ice lattice). This fixes the molecule, prevents water evaporation, and helps protect the protein from radiation damage - even if bonds are broken, the atoms are fixed in place.

http://faculty.washington.edu/lw32/cryoem_home.php

Preparation of cryo-EM Grids

Sample Freezing with a Plunger



Liquid ethane (Liquid N_2 boils on contact)

Cooling speed *v.s.* Ice forms

Ice contamination

(Dubochet et al, Q Rev Biophys, 1988)

When preparing for cryo-EM samples, one important factor is to avoid formation of crystalline ice (right panel). Formation and size of crystal ices is directly associated with the cooling speed (middle panel). The process to avoid big crystal ice and freeze samples with <1nm ice crystals is call **vitrification**, and the ice is called vitreous ice. To achieve in cryo-EM sample preparation, we use a quick-freezing device (plunger) and freeze the sample in **liquid ethane** at -183°C.
Note: Liquid nitrogen, while colder than ethane, is not good, because it boils samples first and causes easy formation of big crystalline ice.

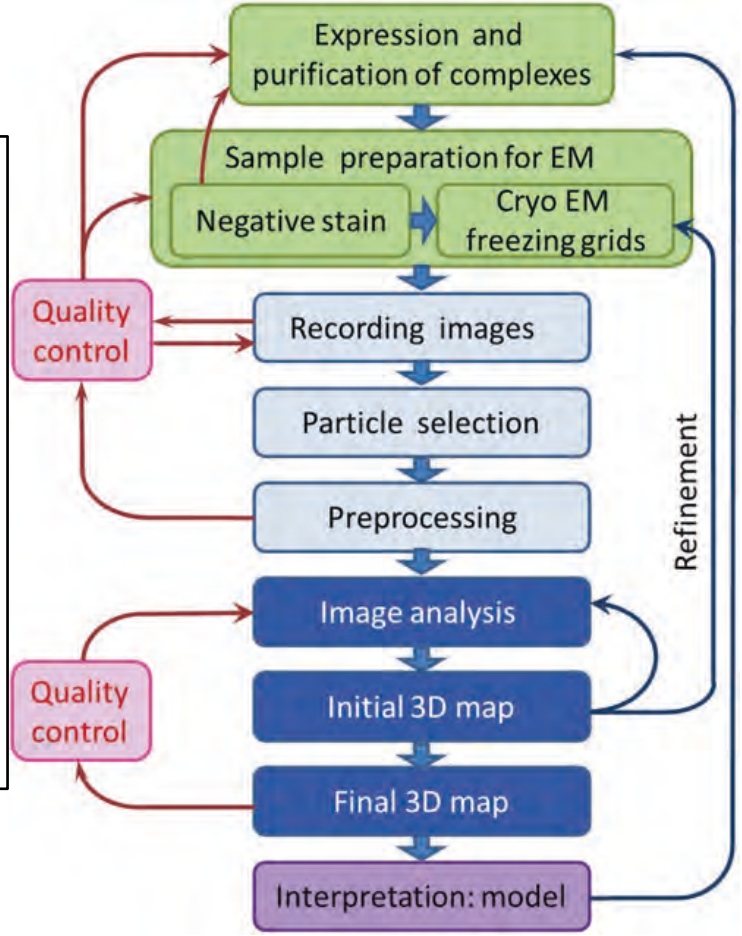
Cryo-EM Workflow

Green:
Protein biochemistry
Sample preparation

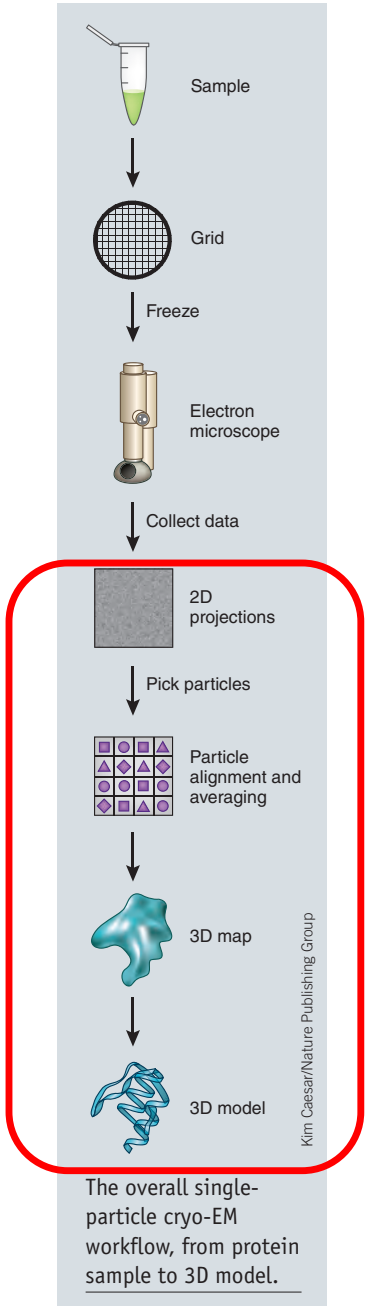
Light blue
Imaging
Data collection

Blue
Image processing
Model building

Purple
Model validation



(Costa et al, Meth Mol Biol, 2017)
(Dörr, Nat Meth, 2016)

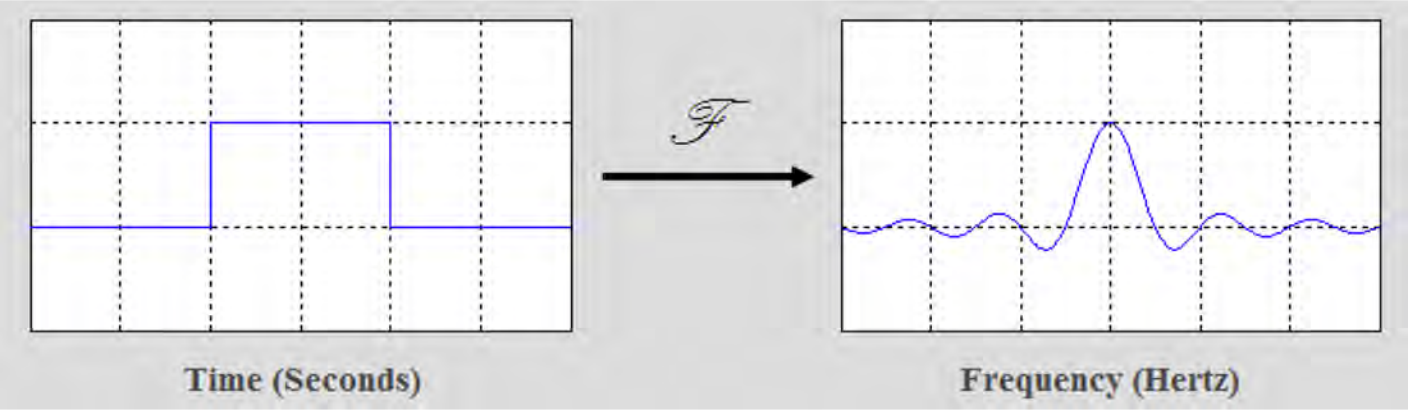


The overall single-particle cryo-EM workflow, from protein sample to 3D model.

Fourier Transform

$$\mathcal{F}\{g(t)\} = G(f) = \int_{-\infty}^{\infty} g(t)e^{-i2\pi ft}\,dt$$

$$\mathcal{F}^{-1}\{G(f)\} = g(t) = \int_{-\infty}^{\infty} G(f)e^{i2\pi ft}\,df$$



The Fourier Transform is a tool that breaks a waveform (a function or signal) into an alternate representation, characterized by sine and cosines. The Fourier Transform shows that any waveform can be re-written as the sum of sinusoidal functions.

2-D images to 3-D Reconstruction

General concept to generate 3-D images from 2-D images:

An electron micrograph shows what it looks like a sheet-like picture when the light sources penetrate the samples. This generate 2-D projections.

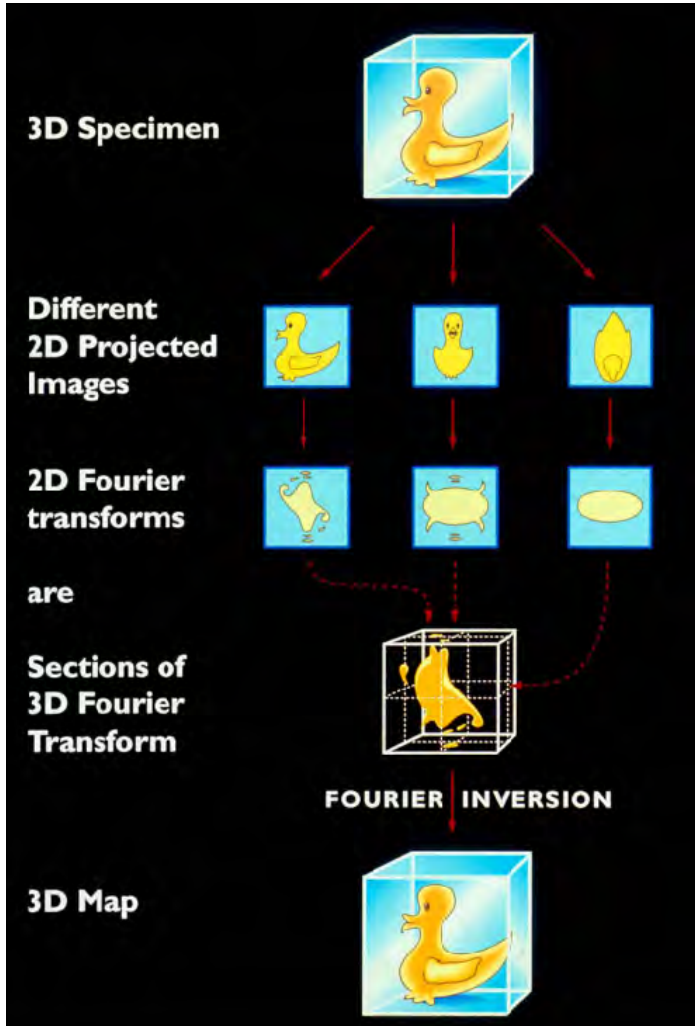
Taking a duck image as an example:

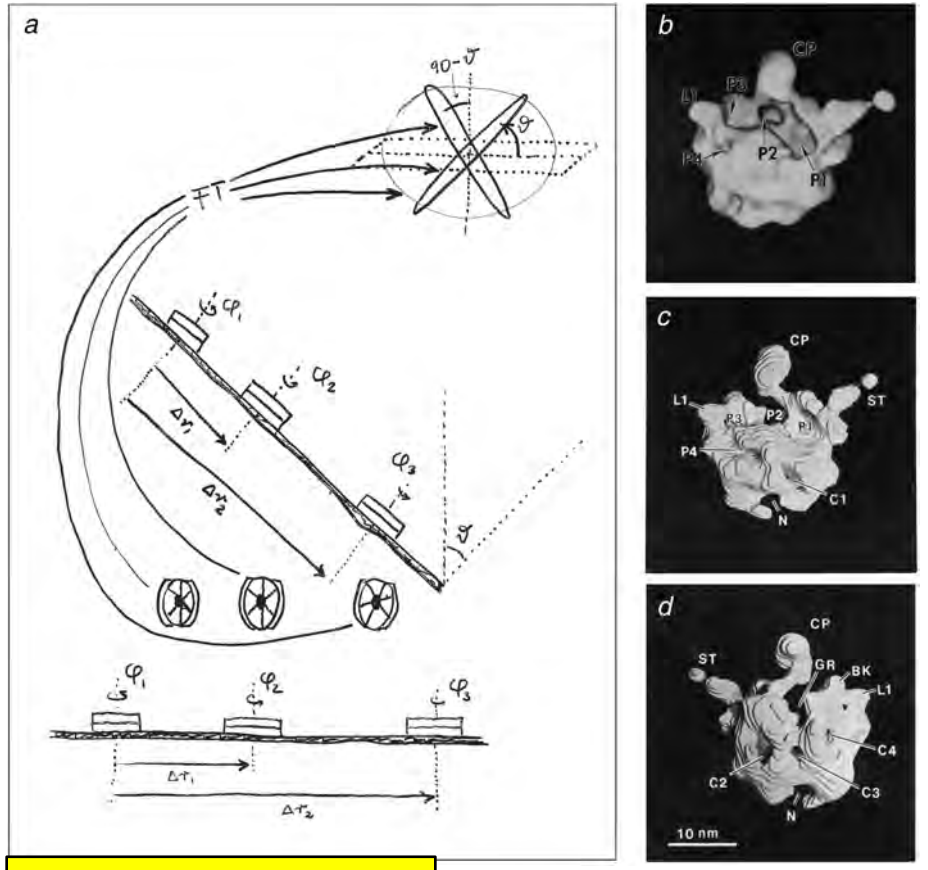
a. One takes pictures of the duck from various angles, where each picture is a 2-D projected image.

b. In order to mathematically combine all 2-D information together, one needs to operate such combination using reciprocal information. Hence converting all 2-D images by FT.

c. Combination of all 2-D reciprocal information results a convoluted image in 3-D.

d. An inverted FT transforms the reciprocal image into a real 3-D image of a duck.





1979 hand-drawing

(Frank, Q Rev Biophysc, 2009)

A. Raw image data typically seen from a cryo-EM experiments. Protein particles are usually shown as the dark objects. This image shows particles of a proteasome complex.

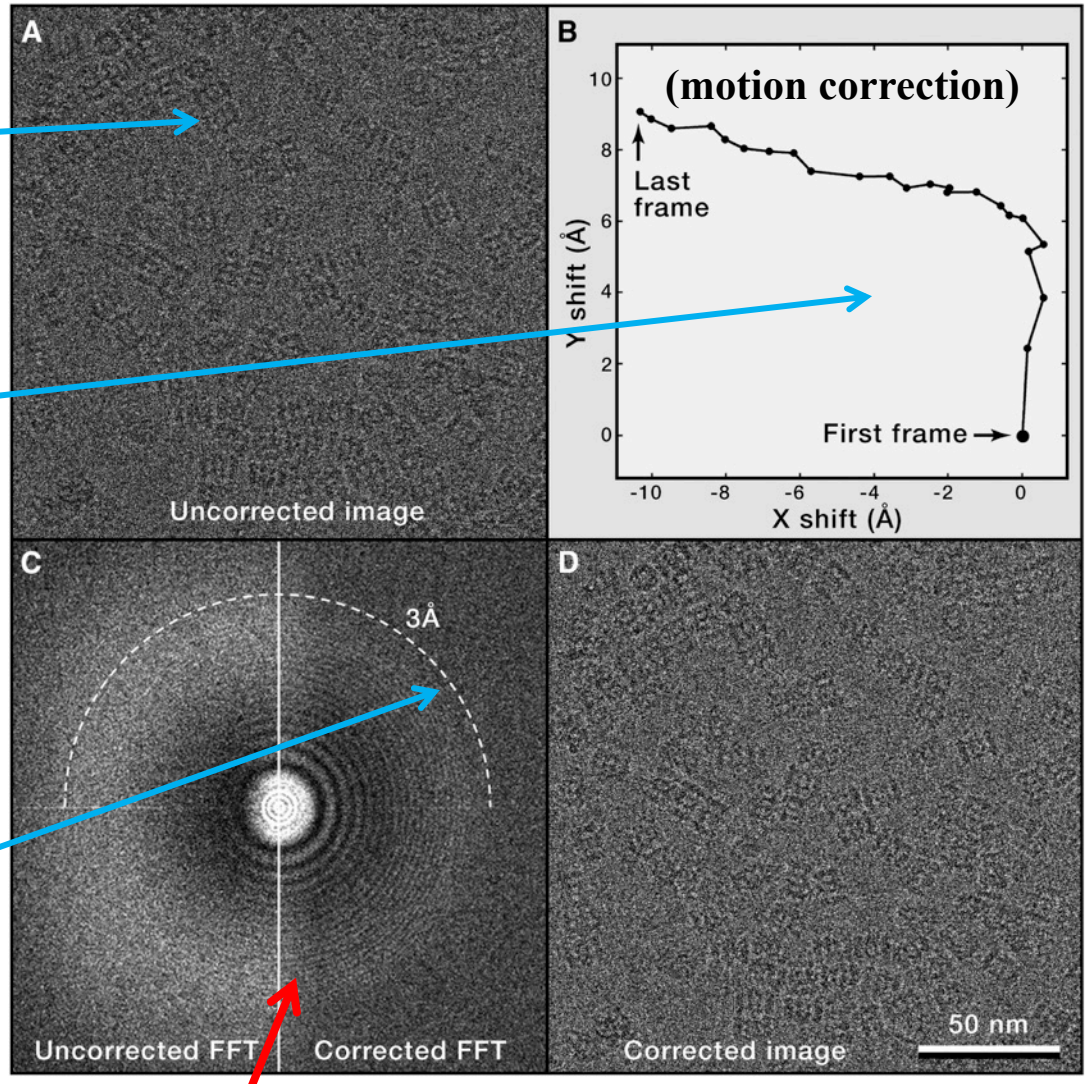
B. However, often time, each particle looks “blurred”, largely because of **image drifting during the picture taking, as well as the microscopic movement of protein molecules in the ice**. ”Motion correction” is thus necessary to help enhance the image quality, i.e., making the images sharper.

C. We use power spectrum to evaluate the quality of an image. Power spectrum can be seen as a theoretical diffraction pattern of the image in A.

D. Corrected and sharpened images from A. As indicated in C, the corrected picture clearly reveals potential information that can be resolved as better as 3Å.

Image film

Power spectrum (Computed diffraction)

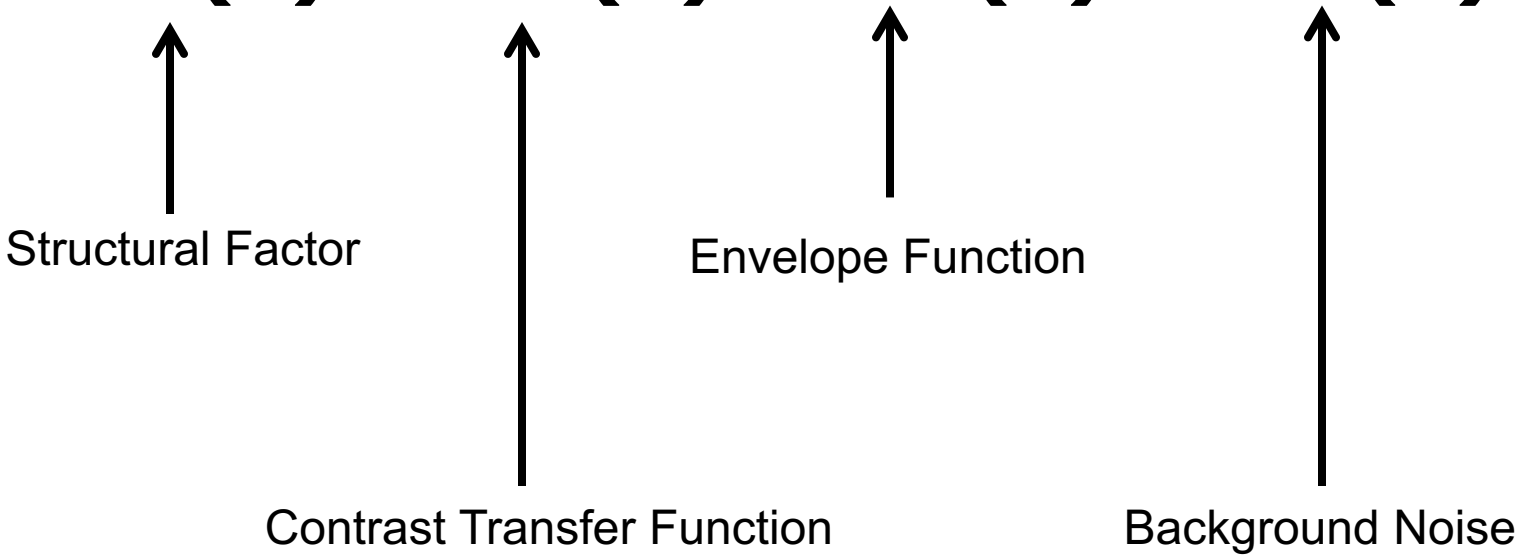


(Cheng et al, Cell, 2015)

What’s this?

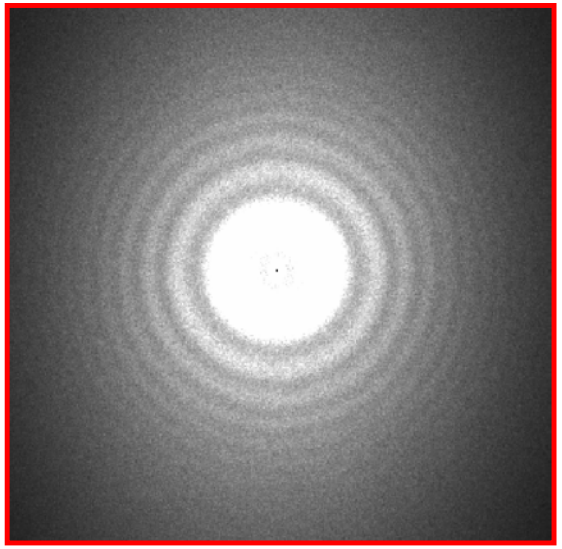
Computed Diffraction Pattern

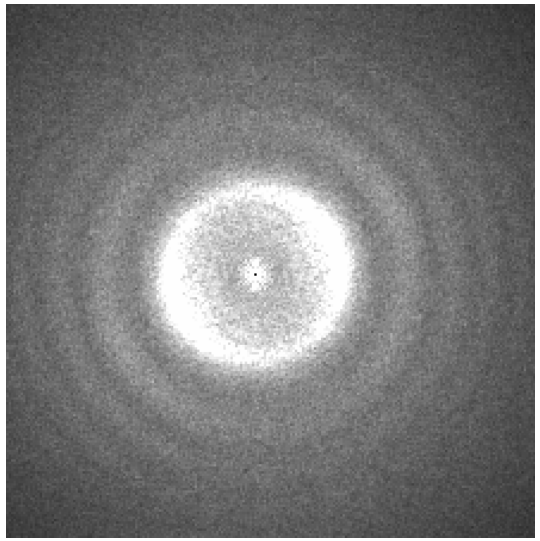
$$F^2(s)\ CTF^2(s)\ Env^2(s) + N^2(s)$$



CTF Assessment (Power Spectra)

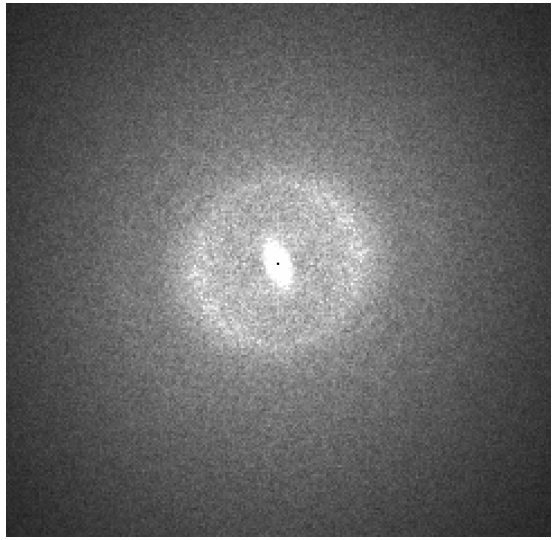
Good:
- Isotropic
- Thon rings at high resolution

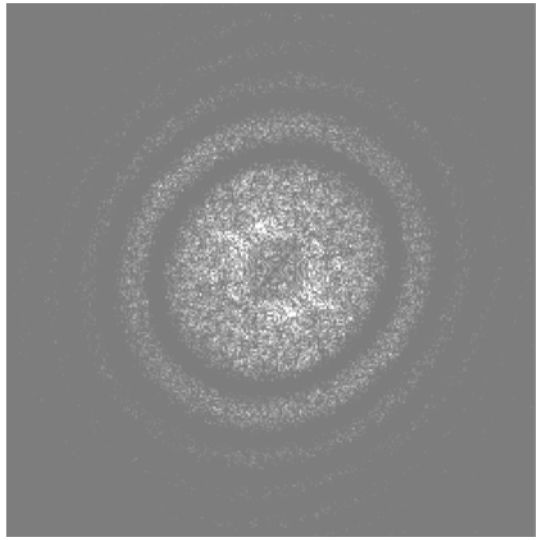




Bad:
Missing Thon rings at certain direction due to drift (*can be corrected if movies are recorded*)

Bad:
Thon rings only at low resolution

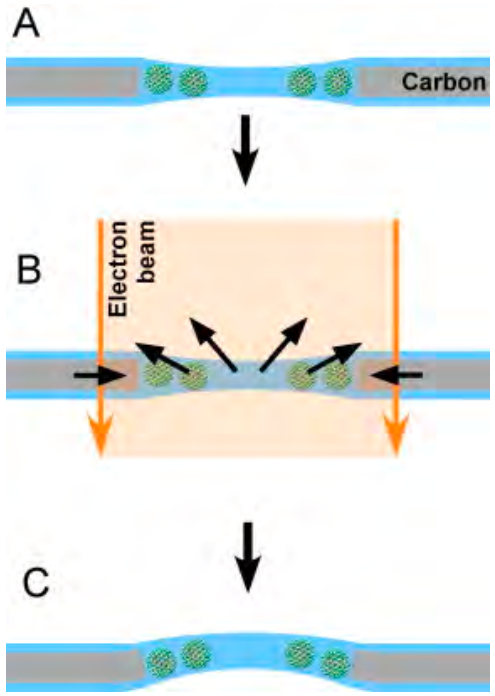




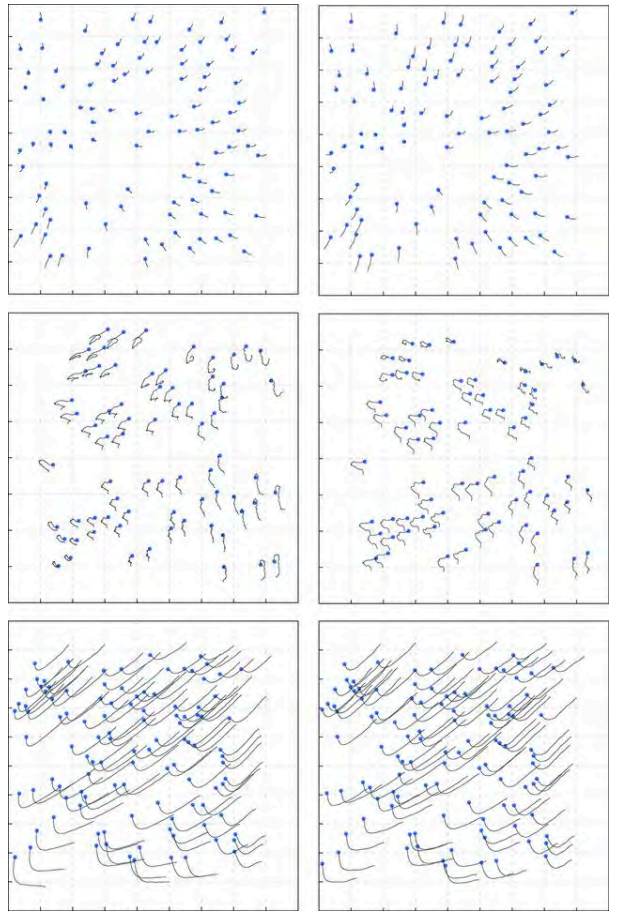
Bad:
Elliptic Thon rings due to astigmatism (*can be useful if properly processed*)

Motion Correction

- Stage drift
- Beam-induced sample motion



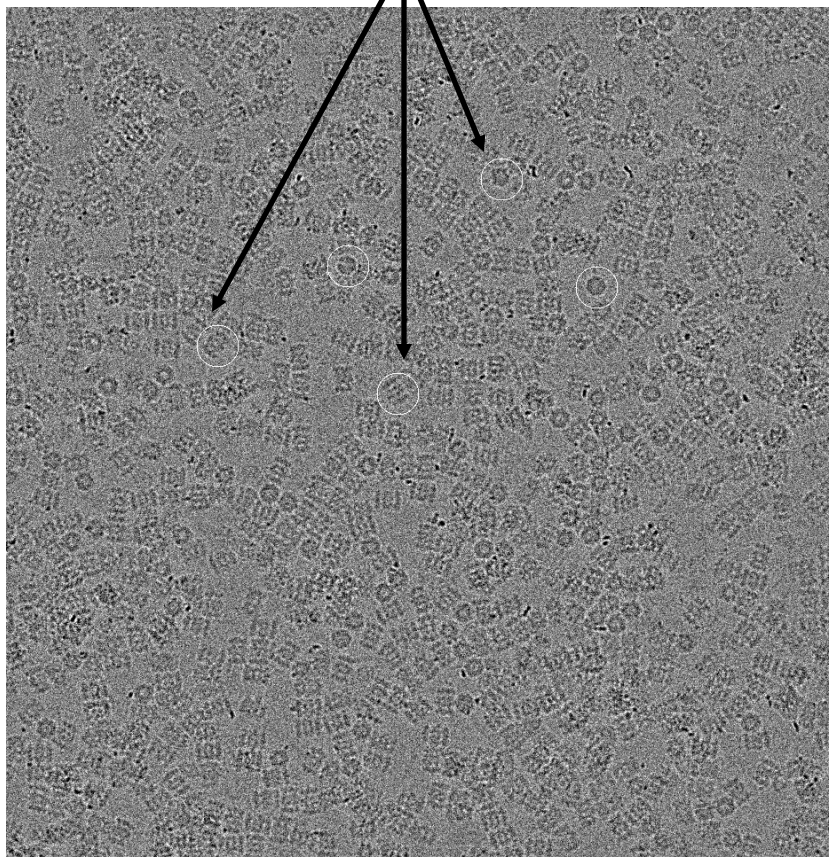
(Brilot et al, J Struct Biol, 2012)



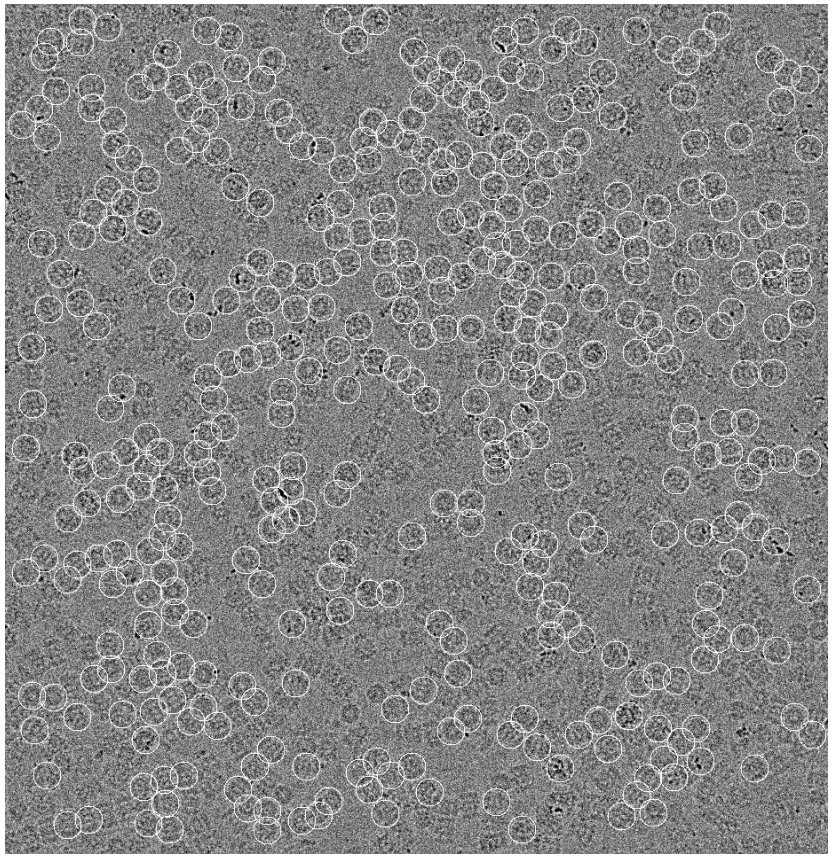
(Zivanov et al, IUCrJ, 2019)

Particle Selection & 2-D Classes

Manual

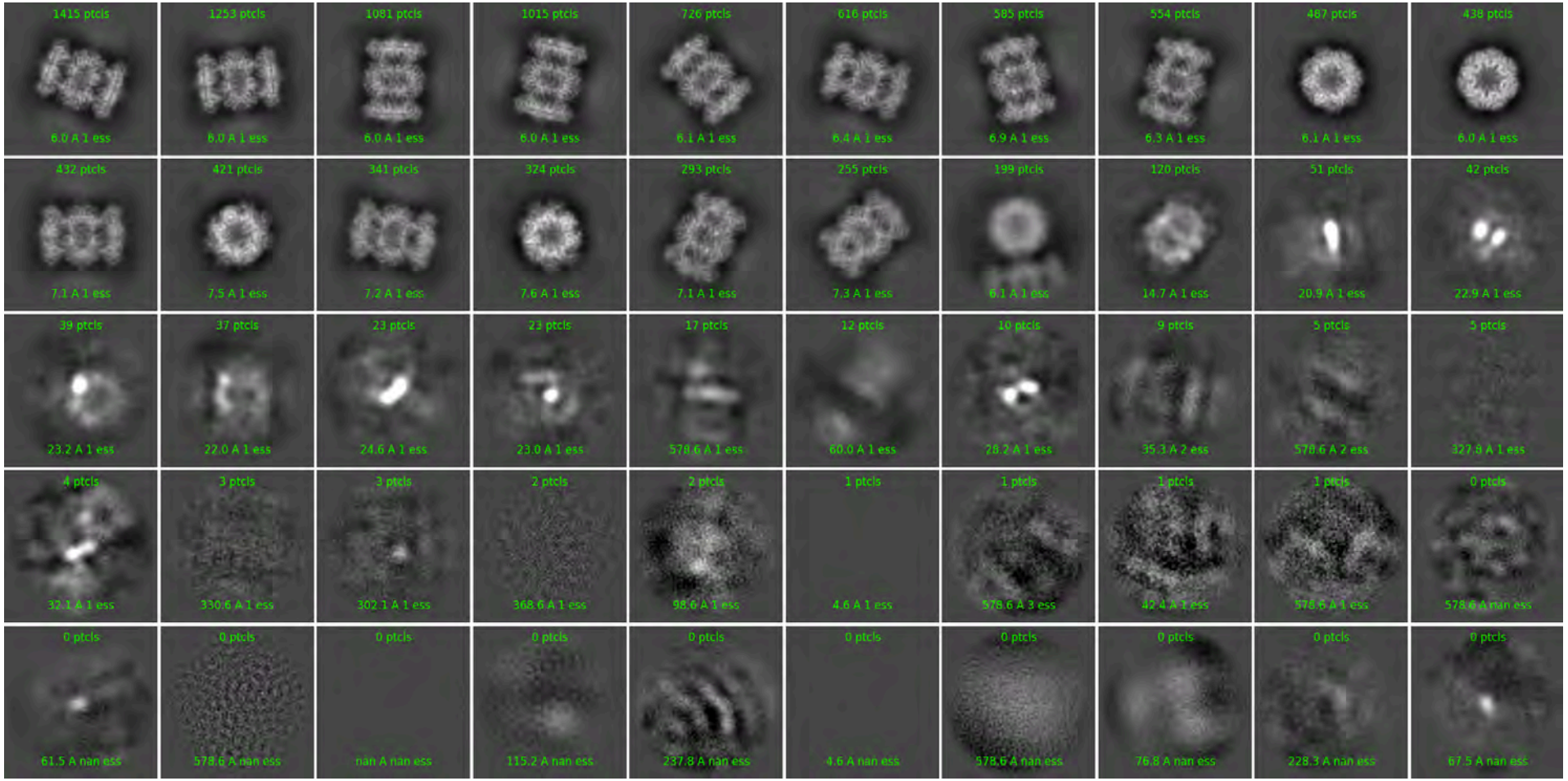


Automated
(template/deep-learning)



Picking particles is the first step of cryo-EM image processing. It allows researchers to generate a library of various shape of 2-D projected images. In general, one can pick by naked eyes (left), but these days, for millions of such particle images, we use some computer algorithm to help pick particle images (right).

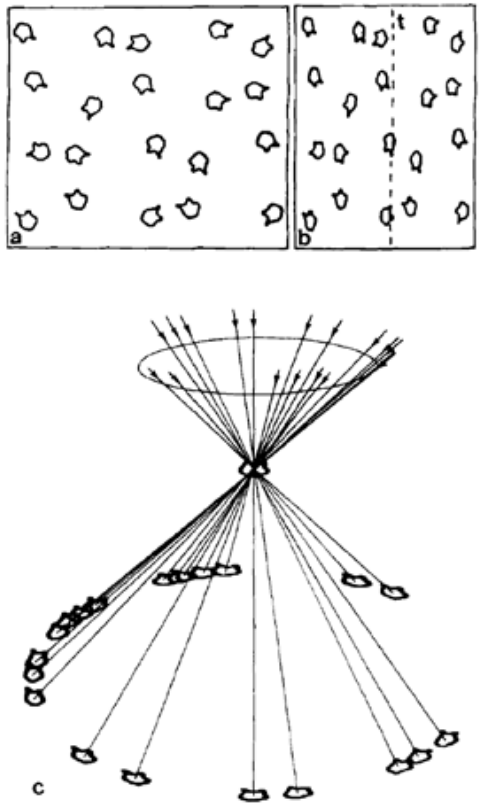
Particle Selection & 2-D Classes



Based on the automated particle selection in the last slide, 50 2-D classes were generated. This process is called ”**2-D classification**”. It shows us what types of 2-D pictures are and whether the automated picking gives us redundant, useful, or useless 2-D images. In the case here, reading from left to right in rows, we see the first 16 classes look reasonable and can be used for further data processing. The rest 34 images either do not have sufficient information or simply garbage.

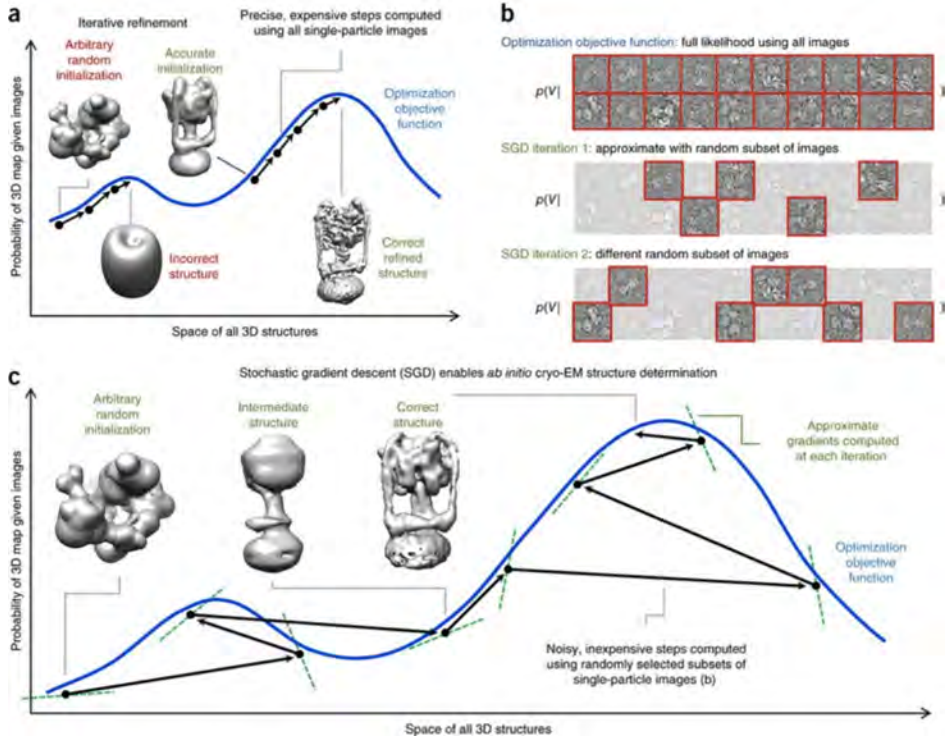
Initial Model & 3-D Classes

Random Conical Tilt



(Radermacher et al, J Microsc, 1986)

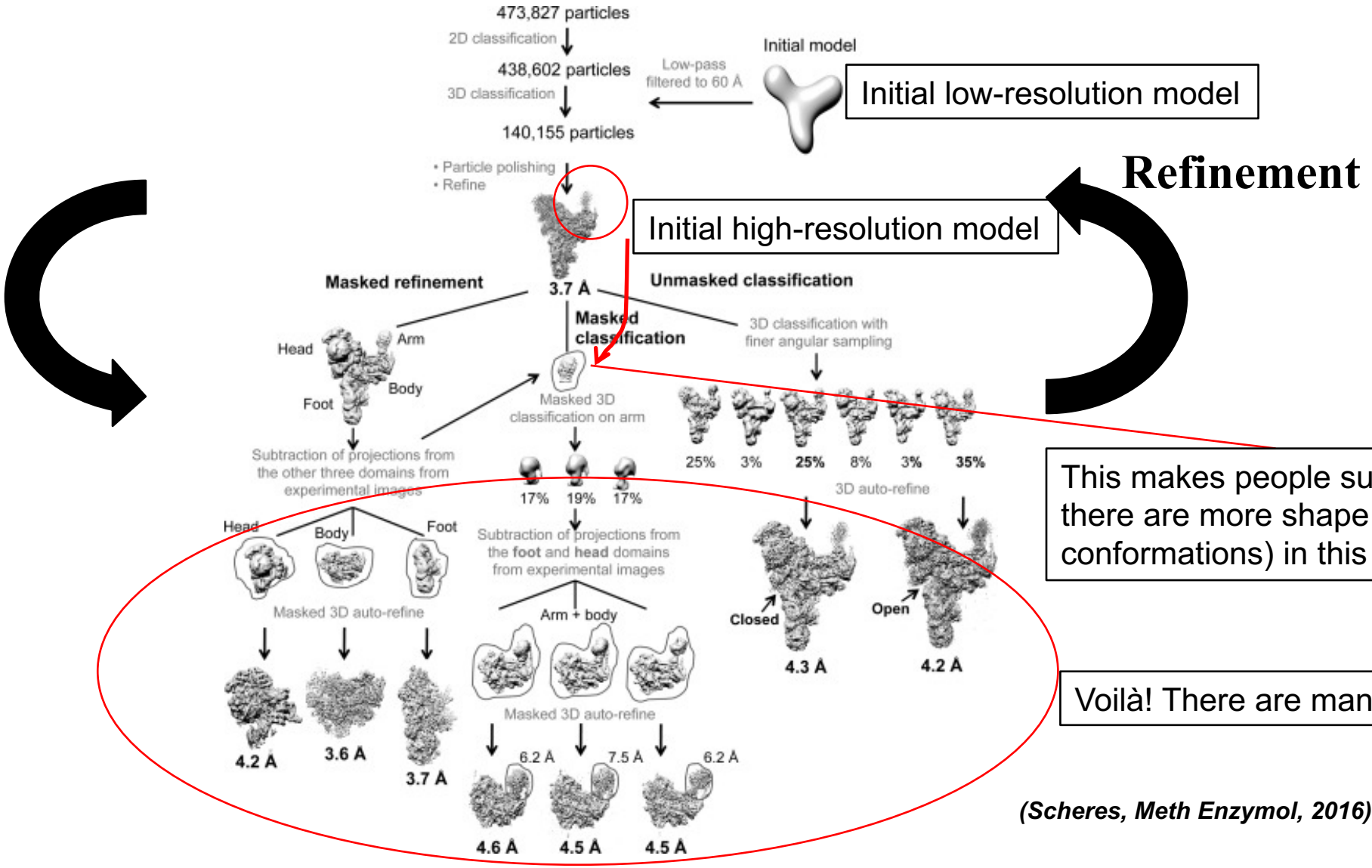
Stochastic Gradient Descent



(Punjani et al, Nat Methods, 2017)

From 2-D images to a 3-D model, there are generally two ways to construct the the initial model and refine the final model. Random conical tilt (left) uses the imaginary angles rendered by each different 2-D projections to rebuild the 3-D picture. Since 2010s, people also started to use machine learning (right) to generate initial model and refine the final models.

Initial Model & 3-D Classes



This makes people suspect that there are more shape changes (i.e., conformations) in this fuzzy area.

Voilà! There are many conformations.

(Scheres, Meth Enzymol, 2016)

Resolution Assessment

- Split particles into even and odd halves, reconstruct and compare models
- Early literature (<2005): differential phase residue
- Now: Fourier Shell Correlation (FSC)
  - Easy to compute. Invariant to filtering/sharpening level
  - Threshold? (0.5, 0.33, 0.14, 3σ)
  - Masking?
  - Split data? When?

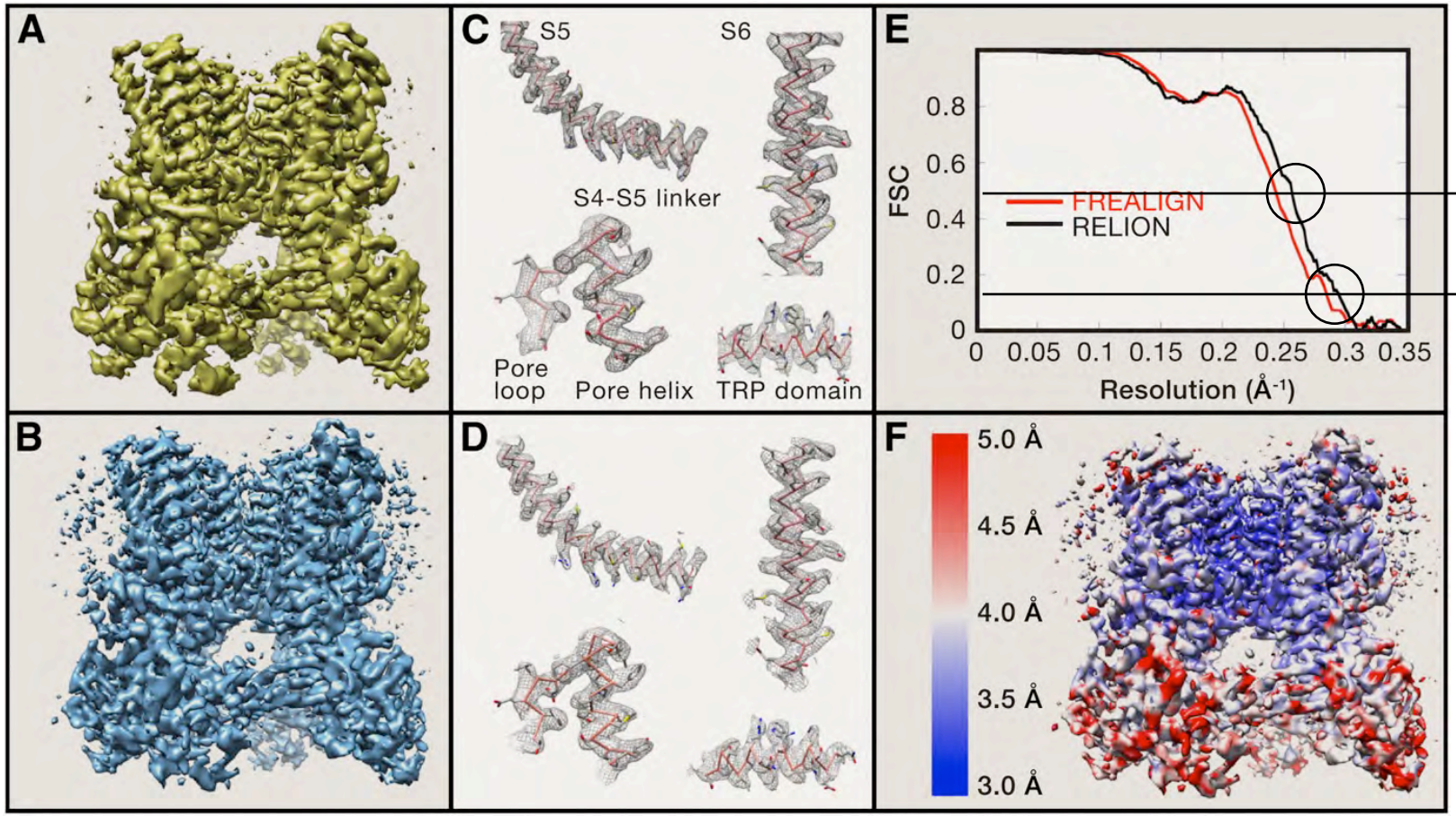
Similar concept to $CC_{1/2}$ as discussed in X-ray crystallography.

$$FSC = \frac{\sum (F_1 \cdot F_2^*)}{\sqrt{(\sum |F_1|^2)(\sum |F_2|^2)}}$$

$$FSC = \frac{SNR}{SNR + 1}$$

Think like % error.

Resolution Assessment: example



FSC = 0.5, Å^{-1} ~ 0.26, ~3.8 Å

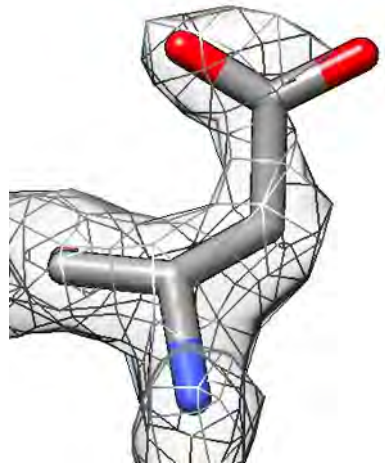
FSC = 0.143, Å^{-1} ~ 0.28, ~3.6 Å

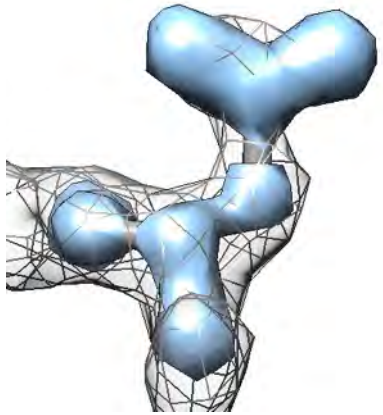
Difference criteria lead to different resolution estimates.

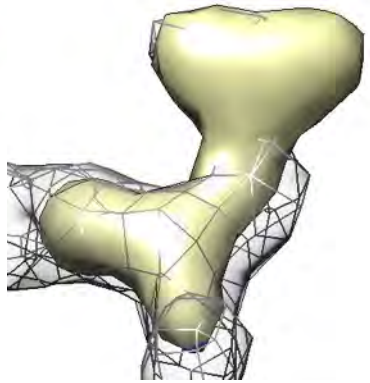
(Cheng et al, Cell, 2015)

Model Validation:

- Average map value at atom positions
- **Map-model Cross-correlation** (in real space)
- Map-model FSC (in reciprocal space)
- EMRinger
- Q-score
- …







Molmap 2Å Molmap 3.5Å

(S²C² Workshop, Stanford)

Model Validation: Model-Map Correlation Coefficient

- Typically referred as CC, map CC, map correlation or real-space correlation

  A metric to show how well the model fits the map.

  (Equivalent to R-factor for crystallography)

Table 3
Summary of map correlation coefficients used in this work.

Metric	Region of the map used in calculation	Purpose
CC_{box}	Whole map	Similarity of maps
CC_{mask}	Jiang & Brünger (1994) mask with a fixed radius	Fit of the atomic centers
CC_{volume}	Mask of points with the highest values in the model map	Fit of the molecular envelope defined by the model map
CC_{peaks}	Mask of points with the highest values in the model and in the target maps	Fit of the strongest peaks in the model and target maps
CC_{vr_mask}	Same as CC_{mask} but atomic radii are variable and function of resolution, atom type and ADP	Fit of the atomic images in the given map

(Afonine et al, Acta Cryst D, 2018)

“Table 1” – data collection and refinement statistics

c. Statistics

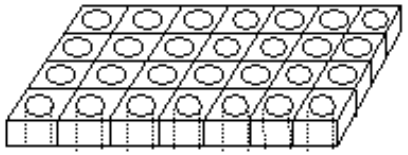
Data collection				
EM	Titan Krios 300kV, FEI Falcon II			
Pixel size (Å)	1.39			
Defocus range (μm)	-0.5 to -3.5			
Reconstruction (RELION)	Overall	Membrane Domain	Peripheral Arm	64k class (for 42kDa)
Accuracy of rotations (°)	0.573	0.711	0.728	0.591
Accuracy of translations (pixel)	0.308	0.400	0.400	0.325
B-factor from post-processing	-88	-85	-83	-89
B-factor for map visualisation	-100	-150	-100	-120
Final resolution (Å)	3.9	4.1	3.9	4.0
Model refinement (PHENIX)	Complete model			
Resolution limit (Å)	3.9			
Number of residues	8037			
Map CC (whole unit cell)	0.758			
Map CC (around atoms)	0.782			
Rmsd (bonds)	0.009			
Rmsd (angles)	1.04			
Average B-factor	86.0			
Validation				
All-atom clashscore	24.4			
Ramachandran plot				
Outliers (%)	0.5			
Allowed (%)	12.5			
Favoured (%)	87.0			
Rotamer outliers (%)	0.1			
Molprobity score	2.5			

Pros and Cons (X-ray v.s. EM)

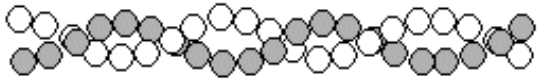
- Pros:
  - Except MicroED/2D crystals, no need of crystallization
  - Smaller demand of purified proteins than X-ray crystallography
- Cons:
  - Molecular size:
    - >200 kD (100-200 kD, pushing the limit)
  - Resolution:
    - Mostly 3-5 Å
    - Overfitting
  - Conformational variability (same issue for X-ray crystallography)
    - Only a small number of functional states are solved.
    - Preferred orientation (single particles in particular)
  - Limited validation criteria

Cryo-EM Friendly Samples

Repeating assemblies

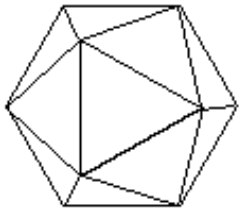


2D crystals
(high tilts needed for 3D)



Helical filaments or tubes
(no tilts needed)

Single particles
no tilting needed,
if randomly oriented



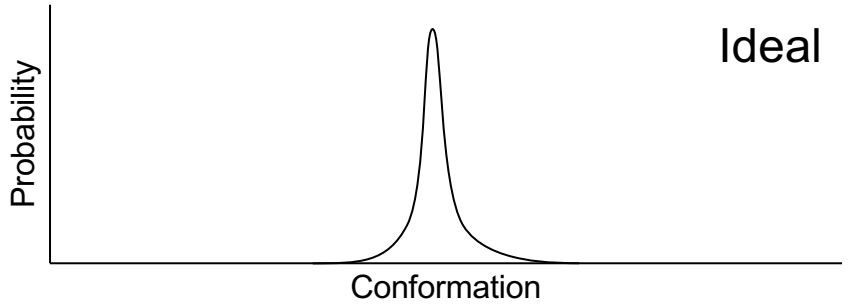
Icosahedral viruses
(or other symmetrical particles)

Asymmetric (>200kDa)
(*e.g.* ribosomes)

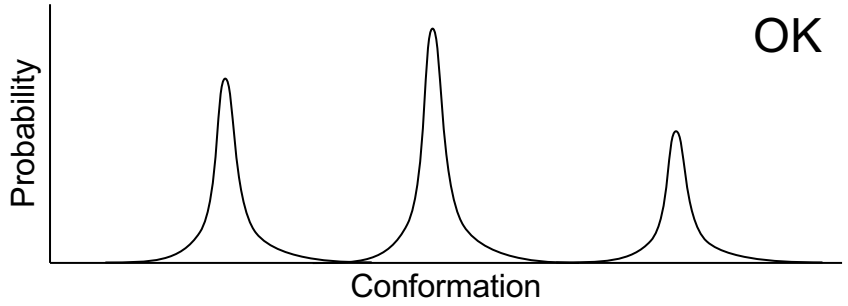
Saibil, Acta Cryst. 2000, D56:1215

Conformational variability

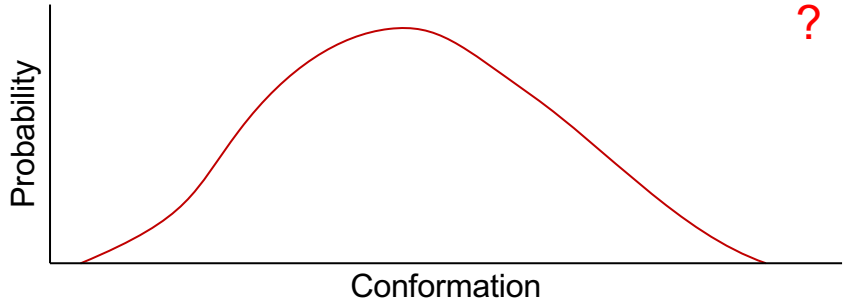
One conformation per image; perfect and simple.

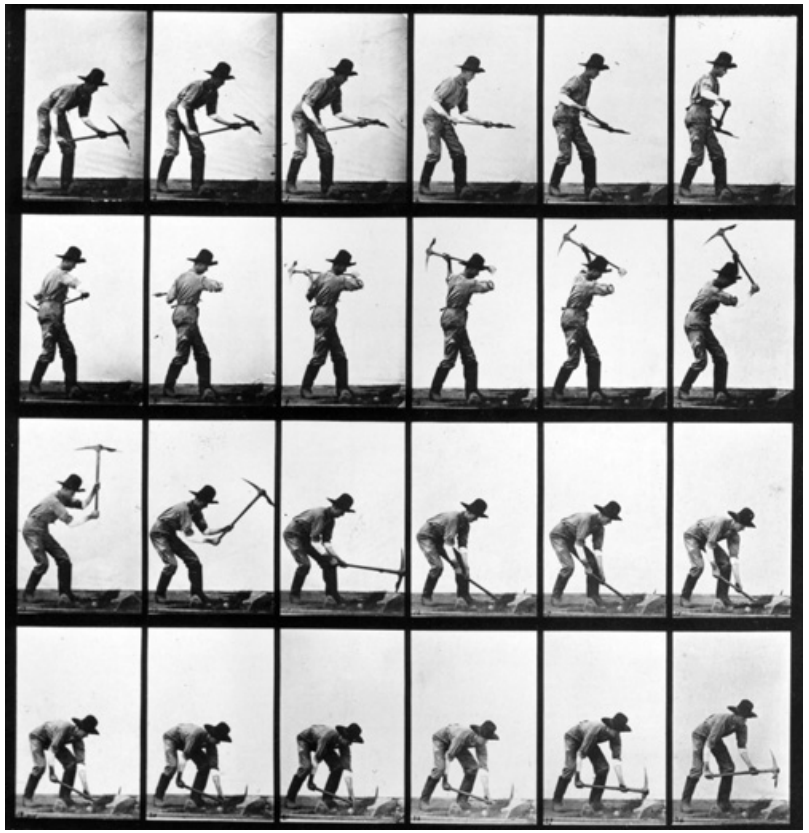


Still OK, if three conformational possibilities per image, then can still sort out the best possible one.



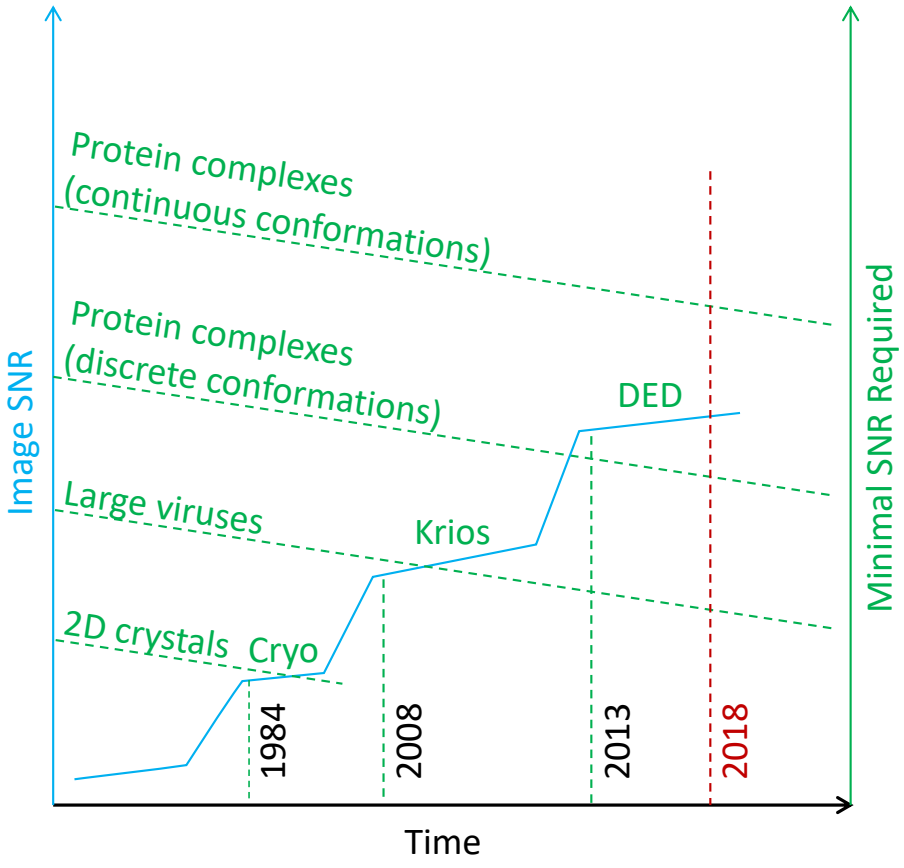
In reality, several conformations mix together for one image, making it hard to tell what specific conformation the image represents.





Only small number of discrete states are solvable now

It’s all about S/N ratio (SNR).



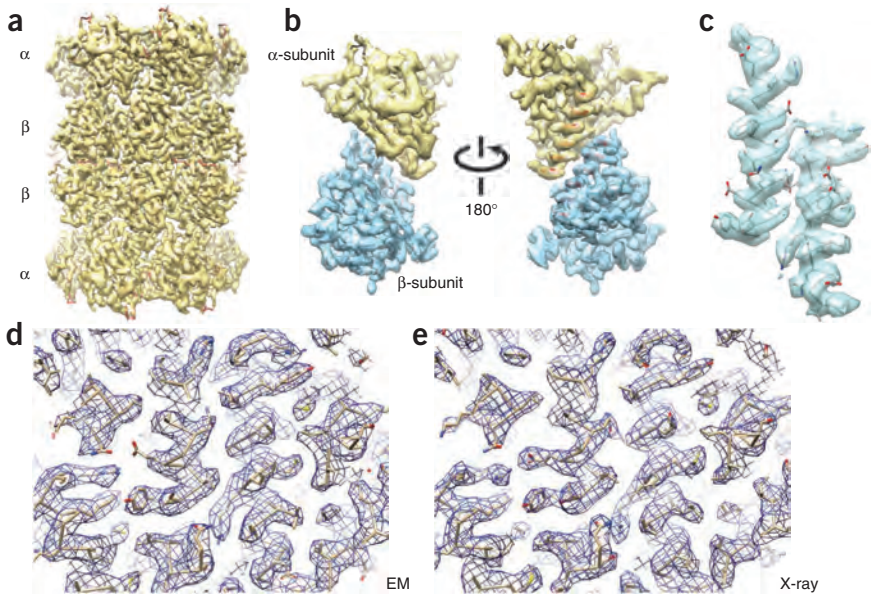
Resolution Revolution?

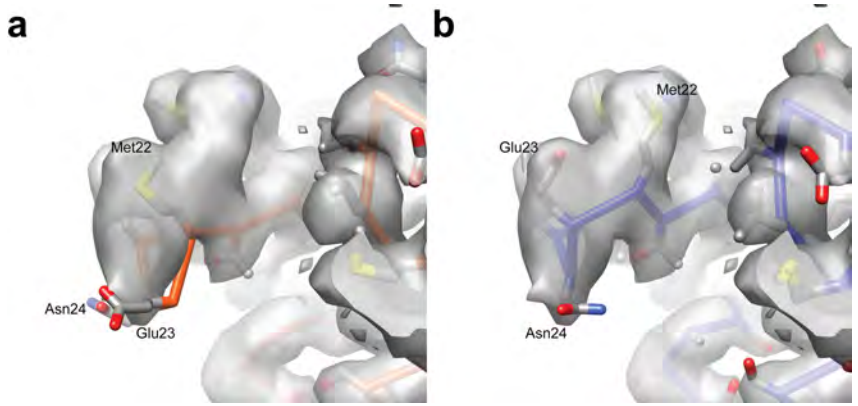
Started 40 years ago.

More to come.

Cryo-EM *v.s.* X-ray Crystallography

Friends and Rivals





a. A short loop (Met22 ~ Asn24) in the β subunit from the crystal structure does not fit well into the final 3D density map, but is easily corrected. While not in a crystal contact, these residues do have much higher temperature factors in the 3.4 Å crystal structure than the average. **b**. We remodeled this loop to fit better into our 3.3 Å density map.

(Li et al, Nat Methods, 2012)